Intoxication, Recklessness and Negligence[i]

By Gideon Yaffe

Introduction

Consider three hypothetical cases: D1, D2 and D3 each takes someone else's suitcase from the carousel at the airport. D1 recognizes a very good chance that it is someone else's but is willing to take the risk; he's reckless in this respect. D2 believes the suitcase is his own but ought to know better. The suitcase looks just like his but as he's in a hurry he does not stop to consider whether he's mistaken, and he ought to have under the circumstances; he's negligent. D3 believes the suitcase is his own because he had several drinks on the plane trip and his intoxication has impaired his judgment. Assume that at the time he takes the suitcase D3 is not so drunk that his bodily movements fail to be voluntary, nor does he meet the law's definition of insanity; in his impaired state he functions about as well as many sober criminals, although less well than he would have had he not been intoxicated. And assume that all three defendants are deserving of some form of criminal liability for taking the suitcase. What is D3 responsible, and deserving of criminal liability, for? There are three possible answers: (1) the same thing as D1, (2) the same thing as D2, or (3) something else—perhaps something a little worse than D2 and a little less bad than D1.

It's a staple of hornbooks that, for the most part, the law accepts answer (1). Typically, intoxicated offenders like D3 are treated as though they were reckless despite the fact that when they acted they were not consciously aware of the risks that their conduct imposed on others.[ii] One way to reach that result, at least to a close approximation, is through the old common law rule, still in effect in many jurisdictions,

that intoxication is to be taken into consideration when identifying the defendant's mental state when the crime is specific intent, but not when it is general intent. General intent crimes are typically equated with those in which the relevant statute is silent about mens rea, and statutes silent about mens rea are typically treated as invoking a recklessness standard. So many crimes of recklessness are general intent. The result is that the defendant who was unaware of risked harms because he was intoxicated cannot appeal to the fact of his intoxication to defend himself from many charges of recklessness. A jury who thought that a sober defendant would have been aware of the risks ought, under this rule, to convict the defendant of recklessness (assuming the crime is general intent). In fact, a jury in such a case may not even hear evidence of the defendant's intoxication, or its impact on his mental state. The *Model Penal Code*, eschewing the fraught distinction between specific and general intent crimes, reaches the same result directly. Under *Model Penal Code* §2.08(2), a defendant who, "due to self-induced intoxication, is unaware of a risk of which he would have been aware had he been sober" is treated as if he were reckless, rather than negligent. D3, then, would be treated as guilty of the same crime as D1.

Interestingly, however, intoxicated negligence is not to be treated as recklessness in the instance in which the defendant became intoxicated involuntarily. If the flight attendant had, for whatever reasons, been feeding D3 vodka tonics while D3 thought, quite reasonably, that he was drinking plain soda water, then he would be not be treated as though he were reckless; in that case, his obliviousness towards the harms his conduct risked would undermine the charge of recklessness. So, for the most part, we find in the law the following principle:

> *The Intoxication Recklessness Principle*: If a defendant is negligent and intoxicated, and he became intoxicated voluntarily, then, for legal purposes, he is to be treated as though he were reckless.

The Intoxication Recklessness Principle does not involve the assertion that the voluntarily intoxicated and negligent defendant is *actually* reckless at the time he is intoxicated. At that time, he is not consciously aware of the risks which he disregards, and so he is not reckless. Rather, the principle tells us that such a defendant can be treated *as if* he were reckless. The principle, that is, licenses the use of a legal fiction.[iii]

Notice that the Intoxication Recklessness Principle applies only if the defendant is properly characterized as negligent before the principle is applied. The principle, that is, tells the jury to treat the defendant as reckless *if he is found first to be negligent*. In fact, intoxicated defendants can be found to be negligent in two different ways. First, the jury might think that a reasonable person *who was impaired in the same way as the defendant* would still have been aware of the risks of which the defendant was oblivious. The jury might think, for instance, that it was unreasonable of D3 to be unaware of the risk the bag was not his *even taking into consideration the fact that he was drunk*. Alternatively, in assessing negligence, the jury compares the defendant not to an *intoxicated* reasonable person but to an *unintoxicated* one. If the unintoxicated reasonable person would have been aware of the risked harms, then the defendant is negligent. In most, and perhaps every jurisdiction in the United States, it is at least permissible for a jury to take the second approach, and some courts positively require it.[iv] The principle operative in assessing the negligence of the voluntarily intoxicated, then, is the following:

> *The Intoxication Negligence Principle*: If a defendant is unaware of a condition and intoxicated, and he became intoxicated voluntarily, then in assessing negligence with respect to that condition he is to be compared to a sober reasonable person.[v]

Combining our two principles, we find that voluntarily intoxicated defendants who are unaware of the harms they are risking are typically treated as reckless, whether or not a reasonable intoxicated person would have been similarly oblivious. A jury that believes that any reasonable person who was as drunk as D3 would have been unaware of the risk that the bag was not his are to hold him to be negligent through application of the Intoxication Negligence Principle, provided that a sober reasonable person would have been aware of the risks. Then, through application of the Intoxication Recklessness Principle, the defendant is to be taken to be reckless for legal purposes. The result is that a defendant whose mental state is not criminal at all, considered independently of the recent history of intoxication that gave rise to it, will be treated as though he were reckless. Since crimes of recklessness are much more severe than crimes of negligence, this result can have enormous repercussions in the form of much greater punishment, most importantly, but in other respects, as well. Under the Supreme Court's decision in *Leocal v. Ashcroft* (543 U.S. 1, 125 S.Ct. 377), for instance, an alien who kills someone while driving drunk can be subject to deportation thanks to the fact that the crime was one of recklessness, while he would not be so subject if the crime were one of negligence (since it would not then be classified as "violent"). Yet he might not be guilty of the crime of recklessness were it not for application of the Intoxication Negligence and Recklessness Principles.

This paper is primarily concerned with the Intoxication Recklessness Principle, although some discussion will be given, also, to the Intoxication Negligence Principle.

The paper identifies a set of conditions under which it is justified to employ the Intoxication Recklessness Principle, even in conjunction with the Intoxication Negligence Principle. When the relevant conditions are met, as we will see, the voluntarily intoxicated negligent defendant is in a mental state that is just as bad as many reckless defendants. The paper, then, defends the law's current use of the Intoxication Recklessness Principle, but with qualifications, for in identifying the conditions in which the principle is justifiably employed conditions are also identified in which it is not.

Section 1 explains some of the challenges that need to be overcome to defend the Intoxication Recklessness Principle, both on its own and in its use in conjunction with the Intoxication Negligence Principle. Section 2 offers a formal model of reckless and negligent decisions, a model that illuminates some central features of recklessness. The model provides an explanation for why, in general, recklessness is worse than negligence. One upshot of the discussion is that many of the defining features of recklessness and negligence play an evidential rather than a substantive role. They are not contributants to culpability themselves, but are, instead, evidence of the presence of the factors that do contribute to culpability. Section 3 uses the model, together with some subsidiary assumptions, to identify a set of conditions under which intoxicated negligence is just as bad as recklessness.[vi] As explained in section 3, the explanation does not extend to cases of involuntary intoxication. So the explanation offered serves as a justification of the Intoxication Recklessness Principle. Section 3 also identifies a series of conditions that have to be met if it is to be appropriate to hold an intoxicated negligent defendant guilty of a crime of recklessness. As we will see, given the rationale for the principle offered here, some of these conditions are met only if certain

empirical claims about intoxication turn out to be true. However, only further empirical work can determine whether they are.

1. Mens Rea Substitution, Prior Fault and Their Interaction

The Intoxication Recklessness Principle is a principle of two different types. The first type might be called "principles of mens rea substitution". Under such principles, a mental state different from that which is required for a crime can serve for guilt for the crime. Or, put equivalently, under such principles the prosecution can establish the prima facie case against the defendant with a showing that the defendant was in a different mental state from that which the statute under which he is charged requires. The general rule, found in *Model Penal Code* §2.02(5), under which mental states "higher" on the purpose-knowledge-recklessness-negligence hierarchy than that required by statute for criminal liability nonetheless suffice for guilt, is an example of a mens rea substitution principle.[vii] The defendant, for instance, who *knew* that his victim did not consent is held guilty of a sexual assault that required only negligence with respect to non-consent. Such a defendant is not negligent—he might even be aware of a condition that a reasonable person would not be aware of—but when it comes to guilt for a crime requiring negligence he is treated as though he were. His knowledge substitutes for the negligence required for guilt. Or, to use another example of a mens rea substitution principle, under transferred intent doctrine in homicide, the intention to kill person A substitutes for the intention to kill the person mistaken by the defendant for person A. The defendant is guilty of the intentional homicide of someone whom he killed, but did not intend to kill, under a rule that allows substitution of the

intention that he did have—the intention to kill person A—for the intention the law requires.[viii]

The Intoxication Recklessness Principle is a principle of mens rea substitution. Intoxicated negligence substitutes for recklessness. In the prosecution of D3 for a crime of recklessness, the prosecution need only show that D3 was intoxicated and negligent. In general, mens rea substitution principles are justified through appeal to an as-bad-as-or-worse theory. The thought is this: a particular mental state is required for a crime because it makes a contribution to the objectionable nature of the agent's conduct in virtue of which the agent is properly punished; it contributes to making the conduct punishable. But since another mental state that is just as bad or worse can make the same, or a greater contribution of this sort, it makes sense to allow punishment when it is present, even if the mental state identified in the relevant statute is not. The conduct in question, when accompanied by the substituting mental state, is just as worthy of punishment (or even more worthy) than it would be were it accompanied by the mental state the law requires. In exchange for lenity towards defendants—in exchange, that is, for defaulting to the most lenient of the reasonable interpretations of statutes—the state is allowed to hold those whose behavior is worse than that for which a criminal penalty is specified, thanks to their mental states, as though their behavior met the definition of the crime. *An adequate justification of the Intoxication Recklessness Principle, then, must explain why it is that intoxicated negligence is just as bad or worse than recklessness.*

The Intoxication Recklessness Principle is also an example of a different kind of principle; it is a "prior fault principle". Under prior fault principles, conditions that would ordinarily shield the defendant from some form of criminal liability fail to do so thanks to the fact that he is to blame for their obtaining. So, for instance, ordinarily a person would be justifed in trespassing in order to prevent the spread of a fire that

would, in the absence of the trespass, burn a great deal of property. But if the defendant lit the fire with the intention of burning the property, and then had a change of heart, he cannot avail himself of the defense of necessity in order to shield himself from liability for trespass, even if by trespassing he prevents the damage that he originally sought to produce. One might have thought that such a person is guilty of attempted arson (assuming that he did not abandon in a way that provides him with a defense on those grounds), but is not guilty of the trespass that he had to commit in order to prevent the damage he was earlier trying to produce. But, instead, he cannot use the necessity defense to shield himself from liability thanks to his prior fault.[ix] Since the Intoxication Recklessness Principle applies only to those who became intoxicated voluntarily, it is a prior fault principle. The rule allows substitution of one mental state for another—of intoxicated negligence for recklessness—*provided* that it was the defendant's own fault that he was intoxicated. The point can be put in a way that preserves the parallel to other prior fault principles: Under the Intoxication Recklessness Principle, a defendant cannot appeal in his defense to his lack of awareness of the risk of harm he imposed if he was unaware of the risk thanks to the fact that he brought his obliviousness on himself by becoming intoxicated. Prior fault removes a shield to criminal liability for recklessness that would otherwise be available to the defendant, namely lack of awareness of the risked harm.

This is not the only example of a mens rea substitution principle that is also a prior fault principle. The law employs such a principle in its treatment of willful ignorance, also, for example. The person who takes steps to prevent himself from knowing that drugs were placed in the secret compartment of his car before he drives across the border—"Don't tell me what you just put in there!", he says—is guilty of *knowingly* importing illegal drugs despite the fact he did not know, at any time, that

there were drugs in the secret compartment.[x] The law allows ignorance to substitute for knowledge provided that the defendant is at fault for being ignorant and positively sought to avoid criminal liability thanks to such ignorance. In the absence of prior fault, ignorance is just ignorance and cannot substitute for knowledge. But where there is prior fault, the substitution is allowed.

It is not transparent why any mens rea substitution principle should apply only when there is prior fault. The mens rea substitution principle is justified on the grounds that the substituting mental state is as bad or worse than the one for which it is to be substituted. But it is hard to see why a mental state should be any less bad when the agent is not at fault for coming to have it (or, equivalently, any worse when he is). Typically, anyway, we think of mental states as contributing to culpability for wrongdoing thanks to what they say about the agent *at the time that he has them*. Even if we have the impulse to reduce the responsibility of those who are *made* bad, we have to acknowledge that they *are* bad. And mental states are relevant to responsibility because of what they contribute to the agent or his conduct being bad. The person who takes something that is not his, believing that it is not his, shows that *at the time of the act* he does not take other people's property rights very seriously—at least, he does not take them seriously enough to outweigh whatever goods he takes himself to get from taking the object. This morally important fact about him is expressed by the combination of his conduct and mental state *regardless of how he came to have the mental state*. If the same behavior, when coupled with some other mental state, is just as bad as the state of knowledge that the object is not his, then it must be because that alternative mental state expresses the same thing (or something worse) about the agent *at the time of action;* it must make the same contribution to his being bad as knowledge would make. If so, that would justify substituting the other mental state for knowledge that the object is

not his. But, still, the causal history of acquisition does not seem relevant to the question. If one mental state is just as bad as another, then it should be just as bad regardless of how it was acquired. At least, so it seems. And notice that, reflective of this point, most mens rea substitution principles are not prior fault principles. This is reflected in the standard rule already mentioned allowing "higher" mental states to substitute for "lower". Knowledge substitutes for negligence, for instance, even if the defendant is not at fault for the acquisition of the knowledge.

Further, there is something problematic about a mens rea substitution principle that is also a prior fault principle when the mental state that is to be substituted *out*—negligence, in the case of the Intoxication Recklessness Principle—is, itself, ascribed to the defendant in part on the grounds of prior fault. Consider what happens when the Intoxication Recklessness Principle is used in conjunction with the Intoxication Negligence Principle, as it often is. Under the Intoxication Negligence Principle, when the defendant is at fault for being intoxicated, we are to compare him in assessing negligence to the sober, reasonable person. He is expected to behave as well as the sober person, given that it was his fault that he was intoxicated. To then go on, through application of the Intoxication Recklessness Principle, to characterize him as reckless on the grounds, in part, that it was his own fault that he was intoxicated would seem to involve double counting. The defendant is found, first, to be negligent rather than simply faultlessly oblivious on the grounds that it is his own fault that he was unaware of the risks. He is then found to be reckless, rather than negligent, on the grounds that, again, it was his own fault that he was unaware of the risks. The same consideration—the fact that it was his fault that he was intoxicated—seems to lift him, first, from no liability to negligence liability, and then, counting it against him again, from negligence liability to recklessness liability. This does not seem fair, somehow.

So, the advocate of a mens rea substitution principle that allows substitution only in cases of prior fault owes an explanation, and such a person especially owes an explanation when the principle in question allows negligence to substitute for recklessness only when there's prior fault. He must explain why it is that the mental state in question (intoxicated negligence, in the case of interest to us) contributes to making the agent's conduct as objectionable as the mental state for which it is substituted (recklessness, in the case of interest to us) *only if* the agent put himself into that mental state, and the explanation must imply that there is no double counting problem. As we will see, significant steps towards the production of such an explanation, for the case of intoxication and recklessness, will be produced here.

2. The Value Structures Underlying Recklessness and Negligence

The partial rationale for the Intoxication Recklessness Principle to be offered in the next section is most easily expressed through a particular model of the value structures of the reckless and negligent actors, structures that underlie and explain their choices to engage in conduct that risks violation of the legally protected interests of others. The model is relevant given the appealing idea that reckless and negligent choices are morally objectionable in virtue of what they demonstrate about the agent's values. The idea can also be put in the terminology favored by the criminal law theorist Peter Westen: the mental states of recklessness and negligence are culpable, are morally significant and contribute to the morally objectionable nature of the agent's act, thanks to what they indicate about the agent's attitude towards the legally protected interests of other people.[xi] As we will see, we are able to uncover some of the formal structure of those attitudes, and how they differ from one another, through consideration of an

appealing model of the value structures that motivate and explain reckless and negligent choices.

Following traditional rational choice theory, let's assume that when an agent chooses an action that risks harm to another, whether or not he is aware of the risk the act imposes, he assigns some positive value to the action. This assumption is not only harmless, but essential to understanding why reckless and negligent acts are culpable. When we hold people responsible for unintentionally harmful acts, we take them to be acting in a way that is expressive of their values, as opposed to acting in some way irrationally contrary to the values they hold. The problem with risky conduct is that the agent chooses the act despite the risks in circumstances in which we think the risks were sufficient to warrant refraining from the act. Such choices tell us something important about the agent only if they are valued by the agent more than the alternative.

Further, the acts of interest to us are acts that have both some kind of effect on the actor and another effect on other people. We can call these, respectively, the result_for_self and the result_for_others, where it is understood that the result in question includes both things that are intrinsic to the act and things that are caused by it. The value the agent assigns to the act is a function of the value he assigns to both the result_for_self and the result_for_others. Call the value of a result for the agent of the act payoff_for_self(result), and the value to others of the result payoff_for_others(result). The value the agent assigns to the act, then, is a function of payoff_for_self(result_for_self) and payoff_for_others(result_for_others).[xii]

Sometimes we learn a great deal of moral significance about an agent merely from consideration of these values, especially if they differ markedly from the actual values of the results for the agent or for others. Someone, for instance, who is simply

miserable when he's a minute late so that he estimates the value to himself of slowly running a stop sign as very, very high, grants too high a value to payoff_for_self(result_for_self). Or, alternatively, a person might grant too low a value to payoff_for_others(result_for_others). He might take it to be much less bad for others to be a minute late than it is for him, for instance. However, as we will see, a great deal of moral significance can be determined *even if the agent gets it roughly right* about the values of these results for himself and for others. Even if he makes accurate assessments of payoff_for_self(result_for_self) and payoff_for_others(result_for_others), he might be culpable for behavior that risks harm to others. So, we will be assuming that his assignments of value to these results are acceptably close to the right values to assign.

The agent may weigh the payoff_for_self(result_for_self) more or less heavily in relation to the payoff_for_others(result_for_others)—he may be selfish or altruistic and in various degrees—and he may think that the results in question are more or less likely to follow on performance of the act. What this suggests is that he assigns value to the act roughly in line with the following equation, and he performs the act thanks to the fact that the value he assigns to it under this equation is positive.

The Value Equation

Value(action) = payoff_for_self(result_for_self)* prob(result_for_self | action) +
*a***s**payoff_for_others(result_for_others)* prob(result_for_others | action)

Several points of clarification about this equation are in order. First, a word about what the various symbols here represent. The probability terms are conditional probabilities; they are the conditional probabilities of the respective results should the agent perform

the act. Parameter *a* represents the degree to which the agent is aware of, or attends to, the result_for_others. The more he is aware of that, the larger the payoff for others figures in the value he assigns to the act. *a* is a value between 0 and 1, inclusive, and could, potentially, be measured experimentally.[xiii] *s* is a social preference weighting. *s* is a measure, that is, of how much weight the agent places on the impact of the act on other people. It is a measure of the degree to which the agent *cares* about the expected impact of his conduct on others. If he grants the same weight to that as he grants to the act's impact on himself, then $s = 1$; if he grants no weight to it, then $s = 0$. All possible values of *s* are possible. If $s = 2$, for instance, then that implies that the agent grants twice as much weight to the impact of his act on others than of its impact on himself. Such might be the case, for instance, when a parent is considering an act that will benefit himself but might harm his child. If $s = .5$, then he cares half as much about the impact of his conduct on others than on himself. If $s = 0$, then no matter how conscious the agent is of the impact of his act on others—no matter how close to 1 *a* is—the second term of the value equation will again be 0. Someone who is utterly selfish, that is, acts as though his conduct had no effect at all on other people, even though he may be aware that it does; he acts the same way as someone who does not notice that his conduct has an effect on others. As we will see, the *s* parameter is particularly crucial to the story to be told here about the kind of culpability involved in reckless and negligent conduct.

The probability terms invoked in the Value Equation should be understood *subjectively* in the following sense: they are relativized to the agent's *action-guiding* beliefs about the probabilities in question. Say that an agent has a coin in his pocket and he has no evidence that the coin is unfair. In making decisions about what to do, he would be guided by the belief that the conditional probability of the coin landing heads

if flipped is ½. So, for the purposes of the Value Equation, that is the relevant conditional probability. It is important to distinguish the agent's action-guiding beliefs about probabilities from both the probabilities in light of better evidence than the agent possesses, and the probabilities that he is disposed *to report* the relevant conditional probabilities as being. To see the first contrast, imagine that, in fact, it is a two headed coin. So long as the agent does not know this, or, even if he does, somehow doesn't take it into consideration in calculating the probability, then for the purposes of the Value Equation the relevant probability term is ½. The relevant conditional probability is 1 relative to a better evidence base, one that includes the fact that the coin is two-headed. But what matters to the Value Equation is not that probability, but the valule the agent himself believes the probability to be.

To see the second contrast, imagine an agent all of whose evidence supports the conclusion that the coin is fair, and who makes decisions under that assumption, but who nonetheless sincerely asserts that the probability that the coin will land heads if flipped is 0. He is just enormously pessimistic. In that case, for the purposes of the Value Equation, the relevant probability term is still ½. What matters, under the Value Equation, is what probability the relevant events are represented as possessing in those beliefs that actually guide his conduct.

So, the Value Equation involves an assumption about the way in which conduct is influenced by representations of probability values. In particular, it involves the assumption that the representation of probability that influences our representations of the values of particular acts might be quite different from the representations that drive our explicit probability assessments, the kind that we report verbally. This is psychologically plausible. A person, for instance, who is twice as averse to reaching into one dark hole than he is to reaching into another is guided in his decisions by a

representation of the probability of being bitten by a spider in the first hole as twice that of the other. But he might still tell you that he believes the chances of being bitten in the two cases to be the same. What *he tells you*, is not *what actually guides his behavior*. But what we care about for purposes of the Value Equation is the latter. For our purposes, then, descriptions such as "the probability the agent assigns to the event" or "believes the event" to have, should be understood to be referring to the mental representations of probabilities that actually guide decisionmaking on the agent's part, rather than those that the agent would report himself to believe the relevant values to be.

It is thanks to this way of understanding the probability terms in the Value Equation that it is possible to distinguish between the *a* parameter and the relevant probabilities. Imagine someone who flips a coin in the air in order to perform a magic trick. It doesn't matter, for the purposes of the trick, whether the coin lands heads or tails, and so he gives no thought to the question of what the conditional probabilities of its landing heads or tails are when he flips it. But, still, he believes that the probabilities in question are both ½. In that case, however, the relevant *a* value is 0; he does not attend at all to the value for others of the coin coming up heads. Or, imagine someone who takes either of two possible outcomes to be equally likely, given his act, but attends fully to both when making his decision about what to do. The relevant *a* values are both 1, but the probabilities he assigns to each of the outcomes is ½. While attention and probability assessment may be linked—one may influence the other—they are nonetheless distinct.

Notice that it is not always possible to determine, from a person's outward behavior, whether he fails to attend to a particular outcome for others ($a = 0$) while assigning it a non-zero probability ($\text{prob}(\text{result_for_others} \mid \text{action}) > 0$) or, instead, takes the probability in question to be 0. In both cases, the second term in the Value

Equation will be equal to 0 and so the impact of his conduct on others will not inform his assessement of the value of the action he performs. This is especially difficult given that people will typically pay no attention to results that they are certain will not come to pass, so when prob(result_for_others | action) = 0, it is typically also the case that $a = 0$. (There are exceptions: the hypochondriac might judge the probability of infection from kissing his great aunt on the cheek to be zero and yet still attend to the possibility when he kisses her.) In addition, some of our verbal descriptions of our own and others' psychological states are ambiguous in this regard. Someone who says, for instance, after taking a bag that did not belong to him, "I thought it was mine!" could be reporting the belief that the probability of the bag's being someone else's was 0. Or he might be reporting the fact that he gave no thought to the possibility that it was not his bag when he took it (although he recognized a non-zero probability that it was not his). We cannot tell which is true from what he says all by itself, even if we assume that he makes a sincere assertion. To tease apart these two kinds of agents, we need to know about facts about their psychologies that are not apparent from their behavior, including some of their verbal behavior. But, and this is the important point, we are in no worse a position here than we are with respect to many mental state discriminations that we make all the time. When a stranger presses the button for the third floor in the elevator, we infer that he intends to get off at the third floor. We do not typically infer that he intends merely to delay the elevator's progress. But either intention is consistent with the small bit of behavior that we observe. We bring other evidence to bear in making our judgment. And similarly in this case. The man who flips the coin that he believes to be fair is rightly attributed with the action-guiding belief that the chance it will land head if flipped is ½. This is so even if *it is possible*, given that he does not care which face of the coin lands up, that he flips the coin while taking that probability to be

0. Other evidence guides our judgment about his mental state, evidence over and above his overt behavior.

The Value Equation is a model. And, like all good models, it matches reality only approximately. It is quite possible, for instance, and compatible with all that is to be said here, that the degree to which a person attends to a risked harm, and the degree to which he cares about the effects of his conduct on others, are not accurately representable on a numerical scale. If so, then to represent the degree to which a person cares about the impact of his conduct on others (the *s* parameter) as .5 or 2 or 0 is at best a guideline to the facts of interest. What is being approximated by such claims is just the same thing as is approximated by phrases in English such as, "He's selfish" or "He's altruistic". There are surely other ways, too, in which the value equation fails to match the psychological phenomena it models. It assumes that the values of the act for oneself and others can be placed on a single scale and added together, for instance, a claim that might be questioned given that often the values in question are very different from one another (a thrill for oneself versus property damage for others, for instance). Still, just as we can learn a great deal about the world from looking at a map, despite the fact that the world is not made of paper, we can learn a great deal about the phenomena of non-intentional harming, and what makes it culpable, from looking at the Value Equation.[xiv]

When it comes to assessing responsibility, we are often particularly interested in the degree to which an agent's act manifested a lack of concern for the legally protected interests of others. Someone who intends harm to another, as an end and not merely as a means, is not indifferent to such interests; he positively seeks to violate them. But when it comes to non-intentional harming, we assume that the agent does not seek to violate other's interests, but, instead, at worst violates them thanks to the fact that he does not care enough about them in comparison to his own interests; at worst, he is not

as concerned with the interests of others as he ought to be. A particularly salient question is *how much* he undervalues others' interests. The more he undervalues such interests, the worse his conduct, and the more deserving of punishment he is for it. Put in the language of our model, when it comes to the assessment of responsibility for non-intentional harming, the value of the *s* parameter is particularly important. It will be assumed here that there is a threshold of concern with the interests of others below which we are willing at least to censure, and possibly to punish. There is some amount of concern with the legally protected interests of others that we take people to *owe* to each other. If a particular agent falls below that threshold, then he is deserving of some form of censure and, possibly, sanction. If this is right, then a crucial question that must be answered in assessing a person's responsibility for non-intentional harming is this: Was the *s* parameter below the acceptable threshold? That is, did the agent weigh the interests of others less heavily than he ought to have weighed them? As we will see, the fact that an agent was in a particular mental state at the time he behaved in a way that harmed another—a mental state that can be characterized in part through an assignment of a value to the agent's *a* parameter—can, together with various other pieces of information, tell us a great deal about how much weight he placed on the violation of the legally protected interests of others.

Before discussing how we can go about estimating the value of the *s* parameter of a particular agent, notice that we can confine our discussion of that issue to those cases in which, relativized to the agent's action-guiding belief about the relevant probabilities, we do not think that the risks were, in fact, worthwhile. The risks in question were not justifiably taken. This restriction is appropriate since we do not assess culpability for taking risks that were worth taking. When the agent's conduct was justified, we do not inquire about the value of the *s* parameter. For our purposes

here, and glossing over complexities, conduct is *un*justified when the following statements are true:

(1) payoff_for_self(result_for_self) > 0

(2) payoff_for_others(result_for_others) < 0

(3) (payoff_for_self(result_for_self) * prob(result_for_self | action) + payoff_for_others(result_for_others) * prob(result_for_others | action)) < j < 0

In other words, an act is unjustified if it promises a benefit to oneself (equation (1)[xv]) and a harm to others (equation (2)), where the expected benefit to oneself is in fact outweighed (no matter what the agent happens to think) by the expected harm to others by some significant amount, *j* (equation (3)). What this suggests is that there are three kinds of actions. There are those that are justified because the potential benefits to ourselves outweigh the potential harms to others. There are those that are unjustified because the benefits to ourselves are outweighed *by a substantial amount* (represented here by *j*) by the potential harms to others. And there are those which we might call "non-justified" in which the potential harms to others outweigh the goods to ourselves but not by much (by less than *j*). The *j* parameter is intended to capture the notion of "gross deviation" from the standard of care of a reasonable person appealed to in definitions of negligence and recklessness. (As we will see, that notion has a somewhat different additional meaning in negligence cases.) It is intended to capture the idea that criminal liability requires more than imposition on others; it requires *serious* imposition of the kind that constitutes wrongdoing warranting the use of state power against one

of its citizens. While the idea of non-justified behavior, in contrast to justified and unjustified, plays little role in explaining what is wrong with reckless and negligent behavior, it will play an important role in understanding the Intoxication Recklessness Principle in section 3.

Compare equation (3) to the Value Equation. The term on the left of equation (3) differs from the term on the right of the Value Equation only in lacking the *a* and *s* parameters. So, given the restriction to cases in which equation (3) is true and given the further assumption that an agent chooses an act only if the value he assigns to it under the Value Equation is positive, it follows that in the cases of interest to us the product of *a* and *s* must be such as to ameliorate the disvalue the agent assigns to the impact of his act on others enough so as to make it outweighed by the value he assigns to the act's impact on himself. That is, it must be because of some mix of the degree to which he is paying attention to the act's impact on others, and the degree to which he weighs that impact, that he chooses the act that imposes risk of harm on others. The combination of dim awareness and insufficient care must result in discounting the disvalue of the impact of conduct on others enough to make it outweighed by the value of the act to oneself.

Given these restrictions, we can make inferences about the value of *s* given information about the value of *a*. In other words, if we know how conscious the agent was of the risked harms, we can make inferences about how heavily he weighed the impact of his act on other people. In particular: *If a is high, then s is low*. The more vividly aware of a risked harm a defendant was when he acted, the more likely he is to have cared so little about the impact of his conduct on others to be worthy of censure and sanction. The more conscious he was of the risks, the more likely he is to have been below the threshold in weighing the interests of others that we take people to owe to

each other. This is the fundamental reason why recklessness is culpable. The mental state of the reckless agent, involving, by definition, a high level of awareness of the risked harm to others, demonstrates the presence of a morally salient fact about himself, namely, how weakly he weighs the interests of others in his calculations about what to do. Since his *a* parameter is high, we are able to infer that his *s* parameter is low. So, his degree of awareness of risked harms is not of intrinsic moral importance. Rather, it is of importance because of what it tells us about that which *is* of intrinsic moral importance, namely the degree to which he cared at the time of action about the negative impact of his conduct on other people.

Figure 1 graphically represents the point just made. Confining ourselves to consideration of cases of unjustified risky behavior, we are able to infer that the agent cares less than he ought to about the impact of his behavior on others, given facts about the degree to which he is aware of harms that he is risking when he acts. If the agent is only dimly, or entirely unaware of the harms that he risks to others when he acts, then we are not able to infer from the fact that the Value Equation is positive that he cares less about others than he ought to. But if he is sufficiently aware of those harms, then the fact that he assigns positive value to the act in accord with the Value Equation allows us to conclude that his *s* parameter is below the threshold of acceptability.

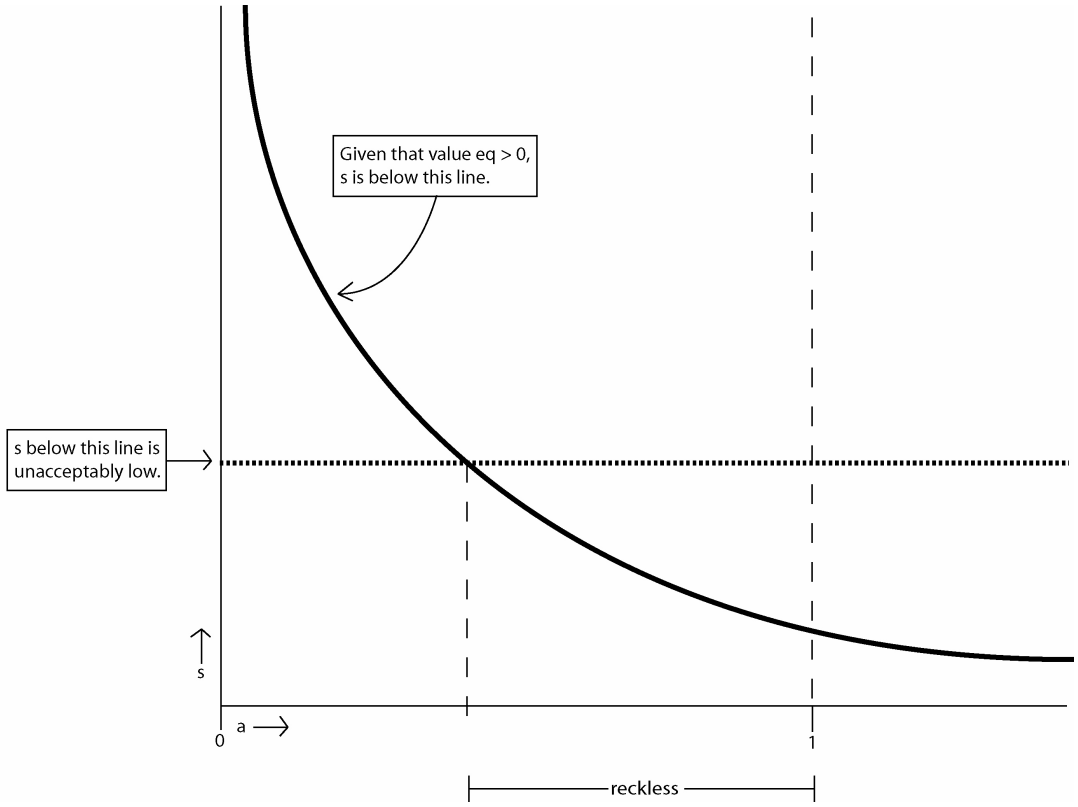

Figure 1: As *a* increases, the maximum value of the agent's *s* parameter decreases, given that the Value Equation is positive. When the *s* parameter is below the threshold of acceptability, the agent is properly characterized as reckless.

This explanation for the culpability involved in recklessness also provides us with the tools for explaining why, with all else equal, it is more culpable to act while knowing that a particular harm will fall on others than it is to do so when there is only a risk of that. If two agents are duplicates in all respects except that one knows his conduct will cause harm and the other is only consciously aware of a risk of that, the former is more culpable than the latter. The difference between the knowing and the reckless agent is not in the value of the *a* parameter, but in the value of the conditional probability of the result_for_others given the act: the reckless agent assigns a value significantly lower than 1 to that while the knowing agent assigns something very close to 1. The knowing agent, that is, is "practically certain" that the result will come to pass. If we assume that all else is equal, then it follows that the knowing agent has a lower *s* value than the reckless agent. If we assume, that is, that the reckless and knowing agent assign the same value to the act, and if we assume that they assign the same value and likelihood to the result_for_self of the act, then it follows that the knowing agent cares less about the impact of his conduct on other people than does the reckless agent. Again, assuming that in cases of non-intentional harming what matters to culpability is the degree to which the agent cares about the impact of his conduct on others—does he care as much as he ought to care?—it follows that there is good reason to take the knowing agent to be more culpable than the reckless.

The conceptualization of reckless choices offered here also helps to explain one of the more puzzling aspects of recklessness. Under the *Model Penal Code*, a choice is not reckless unless the risk is *both* substantial and unjustifiable. The commentary to the code illustrates the distinction between substantiality and unjustifiability with the example of a surgery that has a high probability of death, but in a situation in which

failing to do the surgery has an even higher probability of death. In that case, the risk is substantial but justified.[xvi] What the commentary fails to provide, however, is an example of an act that is unjustified while the risk is insubstantial. By slowing to 10 miles per hour, but not stopping, at a stop sign, an agent can get to work a moment earlier, but imposes a very small risk of killing a pedestrian. We can be virtually certain that the reward is not worth the risk, no matter how small the probability of the harm. Is such an agent guilty of reckless driving? In the context of our model, the claim that the act must be unjustified is the claim that equation (3) is true. But equation (3) could be true even if prob(result_for_others | action) is quite low, below the "substantial" line (whereever that is to be placed). Larry Alexander and Kim Ferzan claim that it would be absurd to acquit a defendant of a crime of recklessness when the act is unjustified merely on the grounds that the probability of the relevant condition being present is not "substantial".[xvii] If the harms to others are significant enough that even a low probability of their occurrence makes them unworth the goods to oneself that the act promises, then the defendant ought not be taking that risk; substantiality is not normatively relevant in itself. Although substantial risks are more likely to be unjustified, they need not be. There is something to what Alexander and Ferzan suggest, but there may still be good reason to deem a defendant reckless only if the probability that his act would cause harm was "substantial".

To see why, note that Alexander and Ferzan measure the degree to which a defendant's act is culpable by the gap between the value the act promises to oneself and the disvalue that it promises to others, corrected for probability. In other words, the more negative the following term, the greater the culpability:

(3*) (payoff_for_self(result_for_self) * prob(result_for_self | action)) +
(payoff_for_others(result_for_others) * prob(result_for_others | action))

Alexander and Ferzan are right to hold this view. After all, what we care about is the degree to which the agent weighs the impact of his conduct on others—the value of the *s* parameter. The more negative the value in (3*), the lower the value of *s* must be (assuming high *a*, as we find in cases of recklessness) in order for the Value Equation to turn out positive. The larger the gap between the value to oneself and the disvalue to others, the less the agent must care about the impact of his conduct on others in order to take the act to be overall worthwhile. However, what Alexander and Ferzan overlook is that the lower prob(result_for_others | action) is, the more sensitive (3*) is to *vacillation* in probability assignment. The result is that, when prob(result_for_others | action) is very small, even tiny mistakes by juries and judges in assessment of the probability of the harmful result for others, can result in big mistakes in their assessment of the defendant's culpability. A numerical example will illustrate the point. Consider the following assignments of values:

payoff_for_self(result_for_self) = 1
prob(result_for_self | action) = 1
payoff_for_others(result_for_others) = -101
prob(result_for_others | action) = .01
a = 1

In such a case, the agent is fully aware of a catastrophic, but very low probability result for others should he perform an act that promises with certainty a small benefit to

himself. The act in this case is unjustified, but it is only very slightly culpable by Alexander and Ferzan's standards. After all, the (3*) term equals 1 * 1 + -101* .01 = -.01. So, the act is not justified (although it may be non-justified, rather than unjustified) but the gap relevant to culpability is quite small. And, as a result, we find that the agent underweighs the effect of his behavior on others only slightly. Under the Value Equation we find the following:

$$\text{Value(action)} = 1 * 1 + 1*s*-101* .01 = 1 - s*1.01$$

Assuming that the agent takes the act to be worthwhile, this value is positive, which implies that $s < 1/1.01$. In other words, the agent will take the act to be worthwhile in this case provided that he weighs the effect on himself only very slightly more heavily—1% more heavily—than he weighs the effect on others.

Now let's imagine that we have mistaken the probability of the harm to another by a single percentage point, although all other values remain the same:

$$\text{prob(result_for_others} \mid \text{action)} = .02$$

Now the (3*) term equals 1 * 1 + -101* .02 = -1.02, rather than -.01. The gap relevant to culpability is more than one hundred times greater than before. And there is also significant increase in the gap between between the value one places on the effect of one's action on oneself in comparison to others:

$$\text{Value(action)} = 1 * 1 + 1*s*-101* .02 = 1 - s*2.02$$

Assuming again that the agent takes the act to be worthwhile, this value is positive, which implies that $s < 1/2.02$. In other words, the agent will take the act to be worthwhile in this case provided that he weighs the effect on himself more than twice as heavily as he weighs the effect on others. If we imagine that the threshold of acceptability is ¾—people are expected to weigh the interests of others at least ¾ as much as they weigh their own—then we find that a very small change in probability, a single percentage point, has resulted in shifting the agent from being well above the threshold of acceptability to well below it.

Notice that we do not have anything like these drastic effects on our assessment when the probabilities of the results for others are quite high. Consider the following assignment of probability, keeping all other values the same:

prob(result_for_others | action)) = .5

Now the act is very culpable. The agent takes there to be a fifty percent chance of a catastrophic result for others should he perform an act that promises with certainty a small benefit to himself. The (3*) term equals -49.5, and applying the value equation, we find that the s value is less than $1/50.5$. In short, the agent who takes the act to be worthwhile in a case like this weighs the value of the act to himself more than 50 times more heavily than the disvalue of the act to others. But say the jury's estimate of the probability of the effect of the defendant's act on others was, again, off by a single percentage point so that, instead:

prob(result_for_others | action)) = .51

Now the (3*) term equals 51.51, and applying the value equation we find that the *s* value is less than 1/51.51. These are small percentage increases in culpability. Morally speaking, there is really no difference between someone who weighs his own interests 50.5 times more heavily than those of others, and someone who weighs them 51.51 times more heavily. These are both people who are rightly described as very, very selfish. Both are well below threshold in their *s* values.

But here's the rub: assessing what an agent takes the probability of a particular result to be is at best a highly approximate exercise. We cannot hope to make such guesses within more than several percentage points of accuracy. In fact, even agents themselves are very poor at determining after the fact what probability assignment guided, prospectively, their decision to act. They are much more comfortable using vague terms like "high" or "low" or "quite likely" than they are comfortable assigning numbers to such probabilities. So, when the probabilities are low—"insubstantial"—there is so much room for inaccuracy in culpability assessment that we ought to refrain from reaching the conclusion that the defendant is worthy of punishment for his risky choices. In short, substantiality matters not because it is of immediate and direct relevance to culpability, but because it is of immediate and direct relevance to *the accuracy of our assessment* of culpability. Only when the probabilities in question are substantial are we able to confidently judge agents to be worthy of punishment for non-intentionally invading the legally protected interests of others. So, contrary to what some commentators have thought, the *Model Penal Code* is quite right to require the risks in question to be "substantial" if the agent is to be taken to be reckless. In taking this approach, the *Model Penal Code* makes a factor that is essential for the adequacy of our evidence of culpability into a defining feature of the culpable mental state. But given that the evidence really is *essential*, this is appropriate.

The kinds of cases that test this conclusion are those in which there is little room to doubt our assessment of the defendant's probability assignment. Imagine, for instance, that the defendant knows that there are 999 blanks in a bag and one live round of ammunition. The defendant pulls a bullet from the bag, loads his gun with it, and fires at the head of a victim. Here, one might think, there is little reason to doubt that the proper probability assignment is .001; the conditional probability of killing the victim given this sequence of acts is 1/1000. But if there is little reason to doubt that, then there is little reason to refrain from concluding that the defendant has the implied unacceptably low *s* value. And yet the probabilities in question cannot be called "substantial". However, even in this kind of case—if it is to be more than a philosopher's play thing—the same issue arises, which, it is suggested, justifies excluding cases involving insubstantial probabilities from the class of the reckless. How confident is the defendant that there were 1000 bullets in the bag, rather than 1001 or 995 or 1005? How confident was the defendant that the gun would fire? Or that the victim would die from the wound? The answers to these questions do not matter much when the probabilities of harm are high. When the probabilities are high, it matters little to our characterization of the defendant's *s* value whether the chance the gun would fire, rather than misfire, was .95 or .90. But when they are low, as in this imagined case, different answers will yield wildly different accounts of the defendant's *s* value. If there are intuitions here that such a defendant ought to be characterized as reckless, they are mistaken. Such a defendant has committed a wildly stupid and destructive prank, but no crime of recklessness. This is not to say that he does not have as low an *s* value as many a reckless agent; he may. It is to say, instead, that we can't be confident that he does.

Explaining why negligence is culpable using the tools on the table requires the introduction of another idea. In particular, it requires noticing that intuitively the *a* and the *s* parameters are not independent. Those who care a lot about the impact of their behavior on others are more likely to be aware of the potential impact of their behavior on others. This is why the doting parent is much more vividly aware of the risks to his child climbing on the jungle gym than he is to the risks to other children who are doing exactly the same thing. When both harms are equally likely—his child is just as likely to fall as the other children are—he is conscious of one potential harm and not the other because he cares about the one much more than he cares about the other. Enough caring brings with it attention. Further, the more pressing the harms risked, the more agents attend to them, other things being equal. So, low level awareness of severe harm to others is a good sign of low weight granted to the impact of one's conduct on others. It is a defeasible sign, however, for there may be other explanations for the lack of awareness in question. Imagine, for instance, that an airline employee mistakenly switched the defendant's name tag from his bag to an identical bag that was not his. When the defendant checks the name tag and reaches the erroneous conclusion that the bag is his, he is not conscious of any risk that taking the bag will invade the legally protected interests of another; he has very low *a*. But the explanation for this fact does not implicate the value that he assigns to *s*. Given the explanation for his unawareness of a risk, we cannot safely infer from the fact that *a* is low that the agent does not place much value on the impact of his behavior on others. We have an alternative explanation available. The employment of the reasonable person standard in negligence is a way of screening out some large number of these alternative explanations for lack of awareness of the risked harms. What it is for a reasonable person to be unaware of a risked harm is for there to be an explanation for the absence

of such awareness compatible with caring a great deal about the welfare of others, as the reasonable person does. In the absence of such explanations for absence of awareness, we can more safely infer that the defendant's *s* value is too low, and so we can more safely infer that he is worthy of punishment when negligent.

Why is it, however, that we think that we are able to infer that a defendant's *s* value is *below the threshold of acceptability* from his failure to attend to the harms his conduct risks? We might grant that a reasonable person would have cared enough to notice, and thus conclude that the defendant cared less than the reasonable person, without concluding that the defendant cared less than he was required to; perhaps the reasonable person cares *more* than is strictly required of him. However, a further plausible assumption allows us to bypass this concern. The assumption is that someone who cared as much as he ought to about the harms his conduct risks to others—someone whose *s* value is at threshold—would be prompted to be sufficiently aware of those harms *so as to recognize that the act is not worth performing*. The reasonable person, that is, does not just notice the potential harms; he notices them vividly enough to lead him to refrain from performing the act from the recognition that the value of the act, as measured by the Value Equation, is negative.

With this idea in mind, a little algebraic manipulation of the Value Equation yields the following. Here *t* is the threshold value of *s*; it is the amount that we expect people to care about the potential harms to others of their acts.

Reasonable person's $a \geq$

(payoff_for_self(result_for_self)* prob(result_for_self | action)) /

-(*t**payoff_for_others(result_for_others)* prob(result_for_others | action))

To illustrate with a numerical example, imagine an act that promises 50 for oneself and negative 100 for others. And imagine that people are required to care about others ¾ as much as they care about themselves; t = .75. It then follows from this equation that the reasonable person's *a* value would be greater than 50/75. More than two thirds of his attention, that is, would be focussed on the potential harm of his act to others. (Attentive readers will notice that the line marking the degree of attention of the reasonable person, by this formula, corresponds to the line demarcating the amount of attention above which the defendant is reckless; see Figure 1.) It is thanks to the fact that a threshold-level of caring prompts this level of awareness or more that the reasonable person refrains from performing acts that are not justified.

As is reflected in the *Model Penal Code*'s definition of negligence (§2.02(2)(d)), a person is negligent not only thanks to the fact that he is unaware of a risked harm of which a reasonable person would have been aware; it must also be the case that the failure to be aware marks a "gross deviation" from the reasonable person. How should we model this idea? What relation need the reasonable person's *a* value bear to the defendant's *a* value for the defendant's failure to be aware of the risked harms to amount to a "gross deviation"? The answer is that the smaller the gap between the reasonable person's *a* value and the defendant's the greater the chance that there is a gross deviation between the two. The reason is that if even a small amount of attention to the risked harms would have prompted someone who cared enough to refrain from the act, then it would not have been much of a burden on the defendant to attend that much. If, by contrast, a reasonable person would have had to attend greatly to the potential harm to others in order to recognize that the act was not worthwhile, then to demand that the defendant do so would be to demand a sacrifice; shifting attention, after all, is a burden.

The points just made are also usefully illustrated by a diagram; see Figure 2. Figure 2 includes a fiction: unlike in Figure 1, where we have an equation (the Value Equation) linking *a* and *s*, here we know only two things: First, the more a person cares, the more he attends; the function linking *a* with *s* is an increasing function. And, when a reasonable person cares as much as he is required to, he also attends sufficiently that the Value Equation comes out negative; that is, the threshold line for *s* crosses the line mapping the function from *a* to *s*. Where Figure 2 includes a thick line that obeys these two constraints, many other possible functions from *a* to *s* are possible. In particular, and importantly, we have no idea how quickly *s* climbs as *a* increases from 0 to the *a* value enjoyed by the reasonable person. For all we know, when $a = 0$, *s* is very close to, but below, threshold. Consistent with what we know, that is, we can say only that someone who is oblivious cares less than he ought to care; we cannot say *how much less* he cares than he ought to. We cannot say how far below threshold he is.

In short, there are two different ways in which attention is linked to caring. It is linked through the Value Equation so that, given values of *a*, we are able to infer something about the agent's *s* value. This is the link that is relevant for assigning criminal liability to reckless agents, as is illustrated in Figure 1. But attention and caring are also linked through a fact about our psychology: that to which we assign value also draws our attention, and in those who care as much as they ought, potential harms draw enough attention to lead such an agent not to take the risk (when it is not justified to do so). Still, the most lenient assumption that is nonetheless consonant with these facts is this: if the agent is entirely unaware of the risked harm, he must be at least some arbitrarily small amount below the threshold of acceptability in the weight that he places on the impact of his behavior on other people. If it would not have been an undue burden on him to be as aware of those potential harms as the reasonable person

would be—if the gap, that is, between the reasonable person's level of awareness and his is small—then he is negligent.

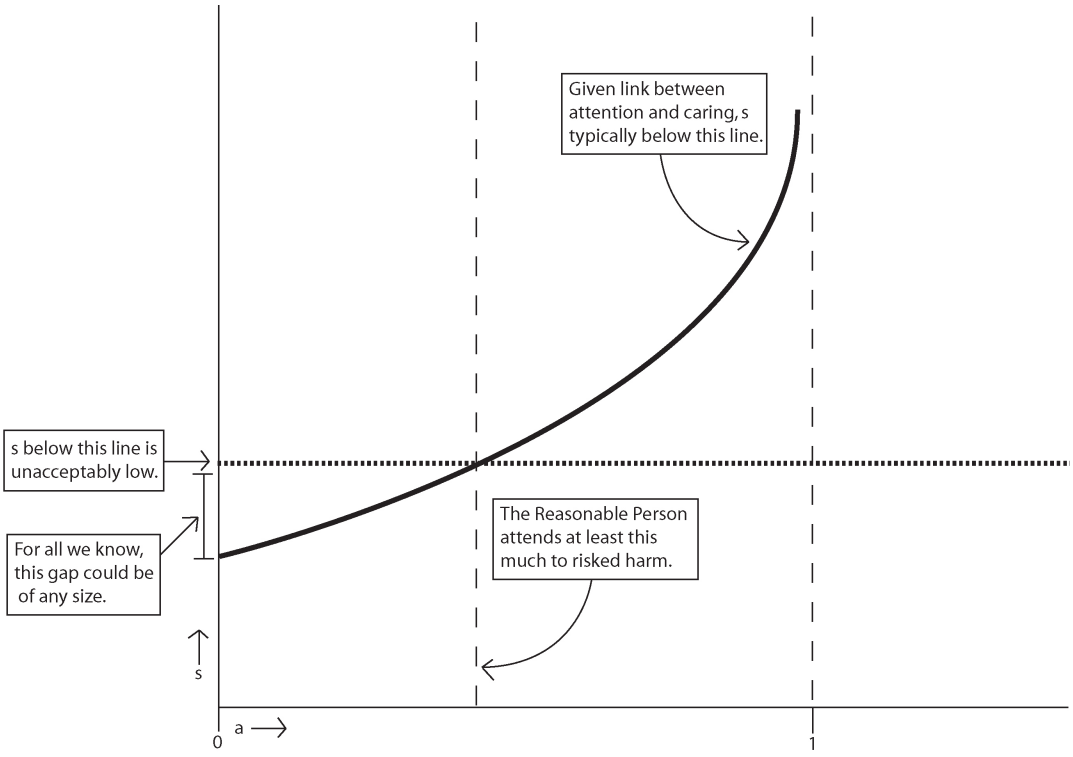

Figure 2: When *s* is at the threshold of acceptability, the agent is aware of the harms that his conduct risks to others to the same degree that the reasonable person would be aware of them. The more he cares about those harms, the more aware of them he is. So, what can be inferred from $a = 0$ is that *s* is below threshold. But since we do not know what the exact function is between *a* and *s*, we do not know how far below threshold a person with $a = 0$ is. If the gap between the reasonable person's *a* level and 0 is sufficiently small, the defendant whose *a* parameter is 0 is negligent.

It is important to note that there are two different, although compatible, conceptions of the reasonable person driving negligence assessments on this model. First, the reasonable person is the person whose level of attention is reflective of the degree to which he cares about others; there is no other explanation for his level of attention. We compare the defendant to this person because if even this person, when oblivious, might have cared about others as much or more than we require people to care, then the defendant is not to be faulted for being oblivious. Second, the reasonable person is the person who, caring as much as he ought about others, recognizes that the act is not justified; the harms it risks are not worth the goods it promises. We compare the defendant to this person once we are convinced that the defendant, had he cared enough, would not have been oblivious. This second comparison allows us to accommodate, in assessing culpability for unintentional harm, not just the burdens and benefits that result from the acts one is considering, but at least one of the burdens associated *with considering them*. This is appropriate since attending is a burden. When only a large attention burden would prompt the reasonable person, who cares as much about others as he ought, to actually refrain from the act, the defendant should not be held criminally liable for failing to.

3. Is Intoxicated Negligence as Bad as Recklessness?

The answer to the title question of this section is "under certain conditions". The burden of the section is to identify those conditions with some precision. Given the model offered in section 2, the question of interest to us can be put like this: Can we infer from the fact that a person is negligent while intoxicated, and became intoxicated voluntarily, that he cares as little about the harms that his behavior risked for others as

a reckless agent in his circumstances would have cared? If, in other words, we are able to infer the same *s* values from intoxicated negligence as we are able to infer from recklessness, then there is good reason to treat the intoxicated negligent actor as though he were reckless, as specified by the Intoxication Recklessness Principle. Using the tools developed in section 2, we are able to identify the conditions under which we are able to make this inference.

To see this, first recall that the line indicating what we can infer about agents's *s* values from their *a* values drawn in Figure 2 concerns the person whose level of attention is just as indicative of his level of caring as the reasonable person whose level of attention is not influenced by anything other than his degree of caring. If we assume that a person's attention is drawn to the harm risked *by* the fact that he cares about that harm, and not by something else—which is part of what we assume when we compare him to the reasonable person—then we can infer from his level of awareness of the risked harm how much he cares about inflicting it on another. So, if we had information about the way in which the defendant deviated from the normal, typical or reasonable person in his *s* value assignments, then we could use that information to justify deviating from the line drawn in Figure 2. Say, for instance, that we know that the agent cares half as much about other people as the normal person does. Then, rather than inferring from the fact that he is unaware of the risked harm that his *s* value is *at* the threshold of acceptability, we can infer, instead, that it is *half way* to that threshold. Notice, and this is a crucial point, that we might infer that an agent cares less about others than the reasonable person cares *from behavior that is not itself punishable since it is not unjustified*. For instance, say that there is evidence that a defendant regularly rolls through stop signs in circumstances in which pedestrians are nearby, but not yet crossing the street. Assuming, as we are, that he makes acceptably accurate

estimates of the value of the possible results of this conduct for himself and for others, from this behavior we are able to conclude that he cares less about the welfare of others than those who stop at stop signs in such conditions. But we are not able to infer that he cares less about others than he ought. It is quite possible that in such cases the benefits to himself, given the very small risks to others, are worth it. So, we know something morally significant about his *s* value—it is lower than that of people who stop at stop signs in these conditions—from behavior that is not criminally culpable.

Now say we had a bit of non-criminal behavior from which we could infer that the agent's *s* value is lower than the norm by some amount x. Later, we find that he is criminally negligent: when he acts in a way that risks invasion of another's legally protected interests, he is unaware of the risk that a reasonable person in his circumstances would have been aware of. Ordinarily, we could infer from his lack of awareness, following the view diagrammed in Figure 2, that his *s* value is some amount below threshold. That, after all, is where we could infer the normal or reasonable person's *s* value to be, given absence of awareness of the risked harm. But, given the earlier behavior, we know that this defendant's *s* value is lower than the normal person's by x. So, we can conclude from his lack of awareness *not* that he is *just below* threshold, but, instead, that he is *x units below threshold*. His negligent conduct is more blameworthy than the negligent actor whom we assume to be normal because his prior conduct allows us to infer from his obliviousness that he cares less about the harms he inflicts on others than the normal negligent actor does.

So, this leads us to the following proposal for justifying the Intoxication Recklessness Principle: from the earlier decision to become intoxicated we are able to infer that the agent cares somewhat less about potential harms to other people than the normal person does; given that, we are able to infer from his later intoxicated

negligence that he cares that same amount less than normal negligent actors care about others; and, further, his failure in this respect is of the same magnitude as at least some reckless actors. The proposal is illustrated in Figure 3.

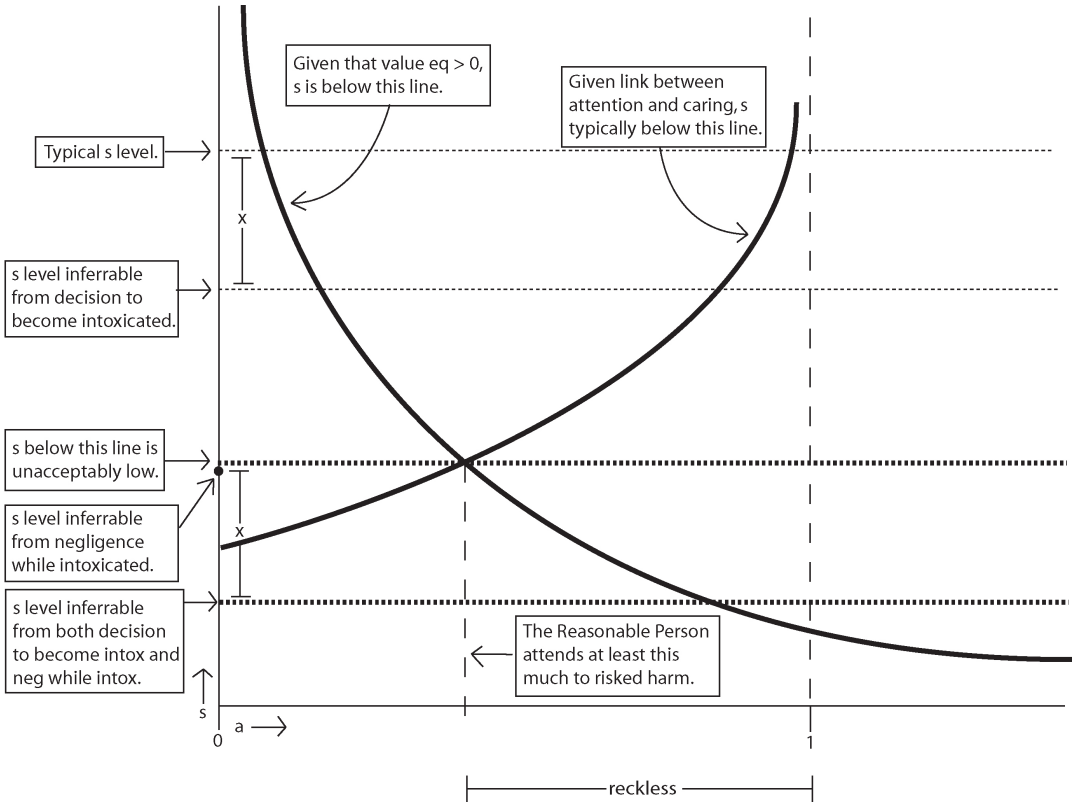

Figure 3: From the decision to become intoxicated, we can infer that the agent's s parameter is below the normal person's by x. From the fact that $a = 0$ at the time of the criminal conduct, together with this first fact, we are able to infer that his s level is x below threshold. This is an s level that is the same as that of some reckless actors.

A pair of large assumptions need to be made for the inference just described to go through. We can infer that the negligent intoxicated agent is as uncaring about the legally protected interests of others as some reckless agents only if both of the following claims are true.

(1) The decision to become intoxicated allows us to infer that the agent's *s* value at that time is lower than the normal person's.

(2) The facts about the difference between the agent's and the normal person's *s* value at the time of the decision to become intoxicated are relevant to the inference from the agent's later negligence to his *s* value at that later time.

In fact, as we will see, both of these claims are false some of the time, and so a court can justifiably appeal to the Intoxication Recklessness Principle in only a restricted range of circumstances, namely those in which both claims are true. However, as we will see, the restricted range is quite large; it's likely that a large majority of cases in which the issue arises are cases in which both assumptions are true.

Consider the first assumption. Under what conditions does a person's decision to become intoxicated allow us to infer that he cares less about the legally protected interests of others than normal, reasonable people do? A first observation is that if the agent is not consciously aware of the risked harm at the time he gets drunk, we cannot infer that he cares any measurable amount less about the interests of others than the reasonable person does. After all, even in cases of negligence, where the conduct in question is not justified, and failure to be aware of the risked harms marks a gross deviation from the reasonable person, we can infer only that the defendant is below

threshold, not that he is any particular amount below threshold. In other words, although we know from his decision to become intoxicated that he cares x units less than the reasonable person about other people's harm, x could be arbitrarily small. So, when, later, we are able to infer from his negligence that his *s* parameter is some arbitrarily small amount below threshold, all we know is that, thanks to his earlier decision to become intoxicated, his *s* parameter is some abitrarily small amount lower, even, than that. Since the sum of two arbitarily small quantities is still arbitrarily small, we cannot infer that his *s* value is as low as that of a reckless agent.

Further, if the act of becoming intoxicated is justified, then, again, the first assumption is false, and there is no reason to think that the agent's intoxicated negligence is as bad as recklessness. Say that someone gets drunk in his living room by himself, expecting to spend the evening drunk in front of the TV before harmlessly passing out. Things might go differently from expected—a friend might come over and, in his intoxicated state, the agent might feed the friend poison thinking it's vodka—but assuming that the chances are low that anyone will interrupt the agent's descent into stupor, and assuming that, even if he were interrupted, the chances are low that the agent would do anything that would injure the other party, then it is quite possible that the agent is fully justified in getting drunk. In a case such as this, in which the act is justified, the decision to become intoxicated is entirely compatible with caring about harm to others as much, or more even, than the reasonable person cares.

Similarly, if the defendant's act of becoming intoxicated is unjustified—if the value of the act to the agent is outweighed by a substantial amount by the potential harm to others—and the defendant is aware of the relevant risked harms when he chooses to become intoxicated, then his act of becoming intoxicated is itself reckless. And as we learned from the discussion in section 2, in cases of recklessness, the

defendant has an *s* parameter some significant amount below the threshold of acceptability. The result is that, in such cases, x is a substantial number, and so we are able to infer from the decision to become intoxicated that the defendant's *s* parameter is not just below that of the reasonable person, but far enough below it to warrant criminal penalities for his behavior.

When we turn to the category of non-justified behavior, things are more complicated. Recall that in these cases the value to the agent is outweighed by the expected harms to others, but not by much—by less than *j*, understood to be the size of the gap needed for the conduct in question to be unjustified. By way of example, consider an agent who tends, when getting drunk at home, to turn up his music very loud for a few minutes, and thereby bother his neighbors. The expected harm to others in light of this—the small harm corrected for the probability that the defendant will actually turn up the music—outweighs the value to the defendant of becoming intoxicated, but by less than value *j*, and so not by enough to warrant characterizing the behavior in question as unjustified. Getting drunk amounts to a peccadillo, we might say, without amounting to a crime, in such a case. Although a reasonable person would not impose such risks on others, the imposition of them does not mark a "gross deviation" from the reasonable person's standard of care. Assuming he's aware of the risked harms when he gets drunk, he has an *s* value lower than the reasonable person's, but not so much lower as to justify criminal liability for his behavior. After all, the Value Equation in his case must be positive, and so he must take the goods of intoxication to be outweighed by expected harms that do not actually outweigh them. This is possible because he can weigh those expected harms less heavily than he weighs goods to himself. So, in his case, there is a gap between his *s* value and that of the reasonable person. He does not care as much about others as the reasonable person

does. This is not to say that he cares so little as to place his *s* value below the threshold of acceptability. We cannot reach that conclusion on the strength of the evidence we have because the gap in value between what his act promises to himself and what it promises to others is insufficiently large to support such a conclusion. Put another way, the reasonable person in non-justified cases has an *s* value that is *j* units above the threshold of acceptability. From the decision to get drunk while aware of the risks, we are able to conclude that the defendant has an *s* value below that of the reasonable person's, but possibly above the threshold of acceptability. But, still, there is a gap between his *s* value and the reasonable person's, and that is all that is required to support the argument for the Intoxication Recklessness Principle on offer here. When, later, he is found to be negligent, he is found to have an *s* value that is at least just below threshold. But given the earlier decision to become intoxicated, we know that it is further below threshold than that. It is less than *j* units *below threshold—x* is less than *j* in non-justified cases—but, still, his *s* value can be inferred to be more than *just* below threshold, and so it is appropriate to treat him as though he were reckless. Many reckless agents, after all, are less than *j* units below threshold in their *s* values.

There are three results of the discussion so far that are worth highlighting. First, since the Intoxication Recklessness Principle leads to miscarriages of justice when applied to people who were fully justifed in becoming intoxicated, or were oblivious of the harms their impairment risked, there is a principled problem in entirely ignoring the culpability of the actor for the decision to get drunk when assessing his responsibility for a crime of recklessness. The intoxicated and negligent offender who takes someone else's bag believing it to be his own should not be treated as though he had been reckless if the chances of his harming others through his impairment were so low as to make his getting drunk worthwhile or if he was entirely unaware of this possible result

of his impairment when he made the decision to get drunk. A similar result follows if intoxication is an anticipated side-effect of medication with medical benefits. If the decision to become intoxicated is worthwhile in light of such considerations, then the fact of its being voluntary should not remove from the defendant the right to shield himself from liability for recklessness through appeal to his lack of awareness. Still, this principled problem is nonetheless a small one, for in the vast majority of cases in which prosecutors are tempted to pursue a guilty verdict, the potential harm to others of impairment are much more important and pressing than the minor, even trivial, goods associated with intoxication. Those who become intoxicated for fun, and later negligently cause harm, are very likely to meet the first of our two conditions under which the argument for the Intoxication Recklessness Principle succeeds.

Second, we have identified a class of cases in which the decision to become intoxicated tells us something of significance about the agent—namely that he cares less about others than the reasonable person does—despite the fact that the agent was not in a criminal mental state at the time he made the decision. These are the cases of non-justified decisions to become intoxicated in which the agent is aware of the risked harms that are imposed on others by his impairment. Since his conduct was non-justified, but not unjustified, in such cases, he is not properly characterized as reckless. His decision deviates, but does not "grossly" deviate, from that of reasonable, law-abiding citizens. But, still, we learn enough about him from his decision to warrant the conclusion that he cares less about others than the reasonable person.

Appreciating the third result at this stage requires noting that in at least some jurisdictions, the person who ingests a substance while negligent about the substance's intoxicating properties is considered to be *voluntarily* intoxicated. (Such is the case, for instance, under *Model Penal Code* §2.08(5)(b).) In the paradigm case, someone offers the

defendant pills at a party without saying what they are. The defendant takes the pills perhaps thinking, unreasonably, that they are candy. Later, when the defendant is intoxicated and negligent—imagine that thanks to his intoxication he fails to be aware of a risk that a sexual partner is non-consenting—he will be treated as though he were reckless through application of the Intoxication Recklessness Principle. Is this the right result? Paul Robinson has answered "no" on the grounds that the defendant is rightly held responsible for the crime of recklessness only if he was *reckless* with respect to the objectionable result of his impairment at the time of his decision to ingest the intoxicating substance.[xviii] As we have seen already, Robinson holds the bar of liability too high: the defendant who is non-justified in his act of becoming intoxicated, but is aware of the potential harms involved in his future impairment, is not reckless when he becomes intoxicated, but we nonetheless learn enough about him from his decision to become so to warrant applying the Intoxication Recklessness Principle to him. But Robinson is right in thinking that there is injustice in taking negligence about the intoxicating properties of the ingested substance to support the judgment that one's intoxicated negligence is as bad as recklessness. To be negligent about the intoxicating properties of a substance one ingests is no different, morally speaking, from being oblivious to the harms one's impairment might cause when one chooses to ingest something one knows to be intoxicating. And as we've seen already, in that case, the Intoxication Recklessness Principle ought not to be employed.

This brings us to the second assumption that needs to be made in order to justifiably substitute recklessness for intoxicated negligence. Recall that assumption:

(2) The facts about the difference between the agent's and the normal person's *s* value at the time of the decision to become intoxicated are relevant to the inference from the agent's later negligence to his *s* value at that later time.

Recall how the proferred rationale for the Intoxication Recklessness Principle proceeds: From the decision to become intoxicated we infer that at that time the agent's *s* value was x units below that of the reasonable person. (So far, we have seen the conditions under which this conclusion is warranted) Then, from the defendant's lack of awareness at the time of action (when he took the bag, for instance) we infer that his *s* value is x units below that of the reasonable, oblivious person. He is, therefore, inferrably x units below threshold, where the normal negligent actor can only be inferred to be just below threshold. He is therefore as uncaring as many reckless actors can be inferred to be, and so he is rightly held responsible as though he were reckless. This line of reasoning fails if we cannot assume that the gap between the *s* values of the defendant and the reasonable person at the time of the decision to become intoxicated is mirrored in a similar gap at the later time of action. That is, assumption (2) must be true if we are to be warranted in substituting recklessness for intoxicated negligence on these grounds.

In general, there are perils associated with any inference about a person's psychology at one time through appeal to facts about his psychology at another, earlier time. People change. So, one set of circumstances in which (2) is false are those in which the agent has had a substantive change of heart between the moment at which he decided to become intoxicated and the later moment when he acts. This is the case in which, at the time of the decision to become intoxicated, he cared less than the reasonable person about the legally protected interests of others, but at the later time of

intoxication he cares as much or more. Given, however, that intoxicants tend to work quickly, and substantive changes in the structure of one's values take time, it seems implausible to think that this happens often through natural means. Still, we can imagine intoxicants that have this effect on people. We can imagine, that is, someone who cares more about the interests of others when intoxicated than he did when he became intoxicated *as a symptom of his intoxication.* Some people become extremely affectionate when drunk. We usually conceptualize this as the alcohol stripping away inhibitions; what is coming out in behavior, are affections that are repressed when the agent is not drunk. But there's no good reason to favor this conceptualization over a competitor: perhaps the alcohol makes the person admire and care about the person at whom he aims his affection more than he did before, albeit only temporarily. We can imagine someone, then, who reacts to intoxicants, or to the particular one he ingested anyway, by becoming *more like* the reasonable person in his *s* value. If the intoxicant has that effect, then (2) is false and we are unjustified in employing the Intoxication Recklessness Principle. In such a case, while we could infer that his *s* value is below threshold from his lack of awareness of the risked harms while in the negligent intoxicated condition, we cannot infer that he is *worse* than negligent, and so we cannot substitute recklessness for his negligent state. It is an empirical question which intoxicants have such an effect on which people.

It is a parallel empirical question which intoxicants have the opposite effect on which people, tending to make them care less about the legally protected interests of others. Anecdotally, cocaine makes one feel special, important, almost invincible. Perhaps, then, cocaine makes one elevate one's own interests over the interests of others. In that case, the gap between the ordinary negligent person's *s* value and the intoxicated negligent person's might be even larger than x, given that it is cocaine that

the agent ingested. This is to say that empirical studies of intoxicants' influence on our *s* values might indicate that we are not only justified in employing the Intoxication Recklessness Principle when certain intoxicants are involved, but could with justice substitute mental states even worse than recklessness for the negligence of the intoxicated defendant.

The second assumption is only true in a yet further restricted set of cases, namely those in which the intoxicant has no distorting influence on the normal way in which caring prompts awareness of risked harms. In order to infer that the typical negligent defendant has an *s* value some unspecified amount below the threshold of acceptability, we assume that had he had a higher *s* value (given the absence of alternative explanations for his lack of awareness) he would have been aware of the harms that his conduct risked. Had he cared more, he would have noticed. But it is not at all difficult to imagine intoxicants that dampen or distort the usual response of attention, or conscious awareness, to social preference. Perhaps it was because he was drunk, and not because he didn't care, that he didn't notice.

It might appear, at first glance at least, that in light of this point the range of cases in which it is unsafe to apply the Intoxication Recklessness Principle is enormous. It is simply obvious that intoxicants dull our perceptual faculties and do so quite often without interfering with our fundamental commitments and valuations. Recall, however, that application of the Intoxication Recklessness Principle yields the conclusion that the defendant is to be treated as though he had been reckless *only if* it is first concluded that he was negligent. That is, we must decide that the defendant *ought* to have been aware of the condition of which he was unaware. In the context of the model introduced in section 2, this amounts to the prior conclusion that his failure to be aware indicates that he has an *s* value below the threshold of acceptability. It is only

once we reach that conclusion that we then apply the Intoxication Recklessness Principle in order to take him to be guilty of the crime of recklessness. The well-taken point that often intoxicants interfere with the link between *s* values and *a* values—with the mechanism, that is, through which caring prompts attention—is to be taken into account in the initial determination that the defendant was negligent at the time of the act. If the defendant was so drunk as to preclude us from reaching any conclusion about his *s* value from the fact that he was unaware of something that a reasonable person would have been aware of, then that fact ought to undermine his guilt not just for a crime of *recklessness*, but for a crime of *negligence* also.

It is relatively obvious that this answers the worry in cases in which the intoxicated defendant would have been aware of the risked harm had he cared a bit more about it. If we find that a reasonable *intoxicated* person, impaired to the same degree as the defendant, would have been aware of the risked harms of which the defendant was oblivious, then it makes good sense to apply the Intoxication Recklessness Principle, even if we think that the intoxicant in question interferes with the usual inference from level of attention to level of caring. We must have gotten past that worry in order to conclude that the reasonable intoxicated person would have been aware of the risked harms. (More, shortly, about the conditions in which it makes sense to have gotten past that concern.) But it is far less clear that this is an adequate response when the defendant is found negligent through application of the Intoxication *Negligence* Principle. Recall that under that principle, a voluntarily intoxicated defendant is found to be negligent by comparing him to the *unintoxicated* reasonable person. That principle leads to miscarriages of justice in cases in which the intoxicant prevents us from inferring anything about the intoxicated person's *s* level from his *a* level. It is only safe to apply the Intoxication Negligence Principle if the decision to

become intoxicated was at least negligent, which is true only if that act was unjustified. In that case, we infer from the defendant's obliviousness at the time of the decision to become intoxicated that his *s* value is below threshold. We then conclude that there are two reasons why, when intoxicated, he was not aware of the harms his later conduct risked: the intoxicant assured that he would not notice those harms even if he cared enough about them, *and he did not in fact care enough about them to notice*. This latter fact justifies our treating him as negligent. But what this implies is that it is just to apply the Intoxication Recklessness Principle in tandem with the Intoxication Negligence Principle only if the decision to become intoxicated was unjustified. It is not appropriate in such cases, if that decision was merely non-justified. The results just reached are summarized in Figure 4.

If the intoxicant disrupts the usual link between a and s, can the Intoxication Recklessness Principle be applied?		
	Intoxicated reasonable person would have been aware of the risked harms	*Sober reasonable person would, but intoxicated reasonable person would not have been aware of the risked harms.*
Decision to become intoxicated is unjustified	Yes	Yes
Decision to become intoxicated is non-justified	Yes	No

Figure 4: There are four cases to consider when answering the highlighted question. Together with the results already reached, the results summarized here indicate that it is appropriate to apply the Intoxication Recklessness Principle in non-justified cases only if a reasonable person who shared the defendant's level of impairment from intoxication would nonetheless have been aware of the risked harms.

In short, then, in cases in which the decision to become intoxicated was non-justified, but not unjustified, we should subjectivize with respect to intoxication when assessing negligence; we should individualize the reasonable person standard in that respect, and before applying the Intoxication Recklessness Principle. That is, in such cases we should compare the defendant to the reasonable *intoxicated* person, contrary to the mandate of the Intoxication Negligence Principle.

It is important to see that in subjectivizing with respect to intoxication (when the decision to become intoxicated is non-justified), we encounter two very different types of cases. There are those in which the intoxication undermines *all* inferences from the defendant's lack of awareness to his *s* value—cases in which it really was just the liquor talking. And there are cases in which the intoxicant's effect is to weaken the response of attention to social preference without eradicating it. In the former set of cases, there is no negligence; we cannot infer that the defendant's *s* value was below threshold; he truly was not himself when intoxicated. But in the latter set of cases there might be negligence given the earlier choice to get drunk which shows his *s* value to be x units below that of the reasonable person. Say that in the absence of intoxication an *s* value of t will prompt awareness of the risked harms; t, then, is the threshold of acceptability. But in the intoxicated, an *s* value of (t + y) is required to prompt awareness; the intoxicant dampens the response of attention to concern with others interests. So, from the fact that the defendant is unaware and intoxicated, we are able to infer that his *s* value is below (t + y), which is y units above threshold. But given that he made the earlier decision to become intoxicated, we are able to infer that his *s* value is x units below the norm. Putting these two points together, we can infer that his *s* value is (t + y – x). So, it appears that we are able to infer that his *s* value is below the threshold of acceptability provided that the intoxicant dampens the effect of caring on attention less

than x units—provided, that is, that $y < x$. Which intoxicants do this to which people? It's an empirical question, not to be settled from the armchair.

A yet further restriction on the set of cases in which it is safe to employ the Intoxication Recklessness Principle emerges from consideration of what Kimberly Ferzan has called "opaque recklessness".[xix] In cases of opaque recklessness, the agent acts while aware of a risked harm under a very broad description, but while unaware of it under the finer description of pertinence to the crime with which he is charged. A person who is texting while driving might be aware that he is risking harm, but unaware that he is risking property damage, much less the particular damage to someone's precious garden gnomes that he ends up causing when he jumps the curb onto the lawn. The question, in such a case, is whether he was reckless with respect to the harm that his act actually caused. The issue arises in this context since, at the time of the decision to become intoxicated, the agent is, typically, aware only that he might be dangerously impaired later, but he is often in no position to anticipate the particular harms that his impairment will risk to others. He is thus aware of the later harms that his later conduct risks only under some broad description such as "harms that I might negligently cause". He is not aware that he might negligently *take someone else's bag*, or negligently *have sex without consent*, or any of the other possible risked harms that are pertinent to the crime with which he is later charged. But, from his earlier decision to become intoxicated, we are able to infer only that he is less caring than the normal person *about the harms that fall under the description of which he is aware, or would have been aware had he cared enough about them to be prompted to think of them*. Since this is different from the description under which he is later to be held reckless under the Intoxication Recklessness Principle, there may be a problem with applying the principle for the reasons proposed.

To make the point more clearly, imagine that we index the *s* parameter to the description of the risked harms under which the agent is aware of them (or would have been aware of them had he cared enough about them). The *s(h)* value, that is, is the degree to which the agent cares about harms *h*. So, from the agent's decision to become intoxicated we are able to infer that his *s(impairment-caused harm)* value is x units lower than the reasonable person's. But when, later, we find him to be negligent with respect to the bag not being his, we've discovered that his *s(taking something not his own)* value is somewhere below the threshold of acceptability. Can we conclude anything about his *s(taking something not his own)* value from his *s(impairment caused harm)* value? Can we conclude that it is x units below the threshold of acceptability? If not, then we ought not to substitute recklessness for negligence in his case. If not, then we are not in a position to conclude that he is as uncaring as the reckless agent about the particular harms that he is charged with having risked. It would be as though we concluded from the fact that someone cares little about his neighbor's interests that he cares little about his child's interests. This is surely a faulty inference; people who care little about one thing often care a lot about others. The question is how much we can learn about the agent's concern with the specific harms his conduct when intoxicated risked from his relative lack of concern with the quite general harms that he was risking by becoming intoxicated.

The general issue here is not at all special to the problem under discussion. When a defendant fails to set his brake before parking on a hill, he risks a large number of potential harms. When his car rolls and kills a particular victim, a stranger to him, it is false that he was aware, or even could have been aware, or that a reasonable person would have been aware, that that particular stranger's life was endangered by his failure. But, still, for many legal purposes we take what he caused to be sufficiently

close to, or "in the scope of", that which he risked (or knew he risked, or would have known he risked had he been reasonable) to warrant holding him responsible for the harm caused. It is not at all clear what principles govern here. It is not clear, that is, under what conditions the risked harms and the caused harms are close enough to warrant liability. But whatever the right principles are, they apply in the case of opaque recklessness involved in becoming intoxicated. If that which the defendant cared less about than the reasonable person at the time of the decision to become intoxicated is close enough to that which the defendant risked through his later conduct while intoxicated, then we can infer that the gap in concern between himself and the reasonable person at the later time is the same as the earlier gap. And if not, then we cannot make such an inference. The driving question is whether his failing to care about what his impairment might cause tells us anything about his failing to care about what his impairment actually risked. The proper characterization of the principles governing the answer to this question are difficult to articulate. But, still, we have discovered a further restriction on the class of cases in which it is justified to apply the Intoxication Recklessness Principle. It is properly applied only if the harm risked by the later action was in the scope of the harms earlier risked by becoming impaired.

In summary, then, we have discovered five conditions that must be met for justice to be served through application of the Intoxication Recklessness Principle:

(i) The defendant must have been consciously aware, at the time of decision to become intoxicated, of the harms that his later decision, while intoxicated, risked.

(ii) The act of becoming intoxicated must have been either unjustified or non-justified.

(iii) The intoxicant must not cause the defendant to care significantly more about the legally protected interests of others than he cared at the time of the decision to become intoxicated.

(iv) The state of intoxication must not undermine the claim that the defendant is negligent while intoxicated. This implies that if the intoxicant disrupts the usual link between caring and attention, the principles summarized in Figure 4 should be followed. This includes subjectivizing with respect to intoxication in the assessment of negligence, if the act of becoming intoxicated was non-justified.

(v) The harms that the defendant's conduct when intoxicated risked must be in the scope of the harms that his decision to become intoxicated risked.

Conclusion

Theorists of criminal law have ordinarily taken there to be only two ways to conceptualize the relevance of voluntary intoxication to culpability. On the first approach, the defendant is assigned responsibility entirely on the basis of his conduct and mental state at the time of the decision to become intoxicated. On this view, when sitting on the plane downing cocktails, D3 was engaged in criminal activity; he was culpably risking taking someone else's bag. Alternatively, one might take the mens rea elements of the crime to have been in place at the time of the decision to become intoxicated and other elements of the crime to have been in place at the later time, when the defendant was intoxicated. On this view, D3 is guilty of the crime of recklessly taking someone else's property because he was reckless when he decided to become

drunk, and he took someone else's property later. On this approach, cases of voluntary intoxication are closely analogous to cases in which a defendant culpably causes an innocent agent to perform a criminal act, as when he gives a lollipop to a small child in exchange for lifting a wallet from someone's purse. The later, intoxicated self is like the small child in the example, and the earlier, unintoxicated self is like the defendant. It is one self who is in the criminal mental state, another who engages in the criminal act.[xx] Under either approach, the Intoxication Recklessness Principle is problematic. Quite often the mental state at the time of intoxication is not one of recklessness with respect to the conditions present at the later time (in having drinks on the plane, D3 is often engaging in non-justified, but not unjustified behavior), and yet the Intoxication Recklessness Principle licenses treating it as though it were.

What has been demonstrated here is that there is a third way of understanding culpability in cases of voluntary intoxication. When coupled with the prior decision to become intoxicated, there are a set of conditions under which negligence is just as bad as recklessness. And when negligence is just as bad recklessness, it is within the bounds of justice to treat the negligent offender as though he were reckless.

This account of the grounds on which the Intoxication Recklessness Principle rests avoids the pair of challenges identified in section 1. First, it explains why it is that prior fault undermines the defendant's right to defend himself from a charge of recklessness through appeal to his lack of awareness of the risked harms: in this case, prior fault shows that the morally salient feature of the defendant that is expressed by his negligence—the degree of weight that he places on the interests of others in deciding what to do—is like that of some reckless offenders. Second, it explains why there is no double counting involved in holding the defendant to be negligent on the grounds that he ought to have been aware of the risked harms of which he was

oblivious, and also holding that he is to be treated as though he were reckless rather than negligent. The solution comes from the fact that his conduct is not promoted from negligence to recklessness on the grounds that he ought to have been aware of the risked harms. The sense in which he ought to have been aware of the risked harms, the sense which justifies taking him to be negligent, is that a reasonable person who cared enough about other's interests would have been so aware. This is not enough to qualify the defendant to be treated as though he were reckless. He deserves to be so treated not because he ought to have been aware but because, given his prior choice to become intoxicated, his failure to be aware tells us more about him than we can learn from his obliviousness by itself. We are able to learn that he cares that much less about others interests than those who did not take steps that caused them to be oblivious later.

The criminal law has tended to treat the state of intoxication as though it were of one of two sorts. Either it is such as to undermine the defendant's capacity for action entirely—and therefore the question arises of whether conviction of the intoxicated defendant is conviction in the absence of a voluntary action—or it is not. In making this gross classification, the law has overlooked the many and varied forms of intoxication and their many and varied implications regarding the morally important features of intoxicated agents that are inferrable from their mental states while intoxicated. All intoxicated voluntary actors are not the same. In identifying the conditions that must be met for application of the Intoxication Recklessness Principle to serve justice, a first small step has been taken for identifying the kind of empirical facts about intoxicants that need to be uncovered if we are to treat intoxicated defendants in a way that matches their culpability for non-intentional wrongdoing. We need to attend not just to the fact that the defendant was intoxicated, nor just to the fact that he became so voluntarily, but also to what we can and cannot learn about the ways in which he

weighs the interests of others from these facts. The framework for conceptualizing the culpability involved in recklessness and negligence offered here helps us to see not just what is important for criminal responsibility for non-intentional harm, but also, in broad strokes, how empirical work on intoxication might usefully inform it.

---

[i] Thanks to Mitch Berman for extremely helpful comments on an earlier draft. Thanks also to Antonio Rangel for many conversations leading to the development of the model offered in section 2.

[ii] See, for instance, Wayne La Fave, *Criminal Law,* St. Paul: West Publishing, 5th ed., 2010, pp. 503-4.

[iii] Most commentators oppose the principle on the grounds that under it defendants who are deserving to be treated as at most negligent (think the commentators) are treated as reckless. See, for instance, Heidi Hurd and Michael Moore, *Punishing the Awkward, the Stupid, the Selfish, and the Weak: Negligence as Criminal Culpability*, Criminal Law and Philosophy, forthcoming, 2011; Larry Alexander, *The Supreme Court, Dr. Jeckyll, and the Due Process of Proof* 1996 SUP. CT. REV. 191, 214 (1996); Paul H. Robinson, *Imputed Criminal Liability*, 93 Yale L.J. 609, 660-3 (1984); Paul H. Robinson, *Causing the Conditions of One's Own Defense: A Study in the Limits of Theory in Criminal Law Doctrine*, 71 VA. L. REV. 1, 16-17 (1985); Kimberly Kessler Ferzan, *Opaque Recklessness*, The Journal of Criminal Law and Criminology, Vol. 91, No. 3 (Spring, 2001), pp. 597-652. One who defends it is Kyron Huigens, *Virtue and Criminal Negligence*, 1 BUFF. CRIM. L. REV. 431, 435-36 (1998).

[iv] For instance, in *People v. Garcia* (250 Cal App.2d 15 (1967)), the voluntarily intoxicated defendant was unaware that the man he was assaulting was a police officer, but was

ruled negligent on the grounds that a reasonable sober person would have been aware at least of the risk of this fact. This approach is typical.

[v] Things are a bit more complicated than this in ways that won't matter for our purposes here. Say a defendant anticipated intoxication, but reasonably anticipated much less impairment than he suffered. Perhaps, for instance, the pills he took were twice the dosage he was told. In such a case, he is not to be compared either to the sober reasonable person, or to the reasonable person who is as impaired as he was. Properly characterizing the state of impairment of the reasonable person with whom he is to be compared, however, is no simple matter.

[vi] In Paul H. Robinson, *Imputed Criminal Liability*, 93 Yale L.J. 609, 660 (1984), a distinction is drawn between justifying the Intoxication Mens Rea PrincipleIntoxication Recklessness Principle on the grounds of substitution and a justification on grounds of accumulation of culpability. Under the first, the mental state of negligence is ignored and the defendant's culpability is assessed through appeal to the mental state at the time he decided to become intoxicated. Under the second, the state of negligence is taken to be as bad as a state of recklessness thanks to the presence, also, of the earlier mental state at the time of the decision to become intoxicated. (Robinson identifies Aristotle as an advocate of something like the accumulation of culpability rationale.) The rationale offered here is of the second variety.

[vii] The idea that this well-known rule of law should be understood as involving a mens rea substitution principle is discussed in Gideon Yaffe, *Attempts,* Oxford: Oxford University Press, 2010, p. 122.

[viii] That transferred intent doctrine should be understood as involving a mens rea substitution principle is defended by Douglas Husak, "Transferred Intent" in *Notre Dame Journal of Law, Ethics and Public Policy*, v. 10, 1996, pp. 65-97.

[ix] Cff. *Model Penal Code* §3.02(2). Notice that this provision does not allow the defendant who is at fault for the fact that he faces a situation in which he must break the law to prevent the greater evil to use the necessity defense against a charge of intentionally or knowingly committing a crime. But he could still use the defense against a charge of recklessness or negligence. Still, there is a form of criminal liability that he cannot defend himself against thanks to prior fault.

[x] The lead case on this issue is *United States v. Jewell* (532 F.2d 697 (9th Cir. 1976))

[xi] Peter Westen, *An Attitudinal Theory of Excuse,* Law and Philosophy, v. 25, 2006, pp. 289-375.

[xii] Given how result_for_others and result_for_self are defined, we can ignore payoff_for_others(result_for_self) and payoff_for_self(result_for_others). If, for instance, the same effect of one's act on oneself and on others, then it will be included both in the result_for_self and result_for_others. So, payoff_for_others(result_for_self) and payoff_for_self(result_for_others) are already incorporated into payoff_for_self(result_for_self) and payoff_for_others(result_for_others).

[xiii] It is assumed that the agent is fully aware of the result_for_self. This is why no separate *a* parameter accompanies the first term on the right side of the Value Equation.

[xiv] For an instructive discussion of the ways in which false claims that nonetheless capture something of the truth are useful for figuring out things about the world, see Elijah Millgram, *Hard Truths*, Oxford: Oxford University Press, 2009. I claim only that

the Value Equation is such a proposition, not that it is what Millgram calls "a hard truth".

[xv] There are also unjustified behaviors that promise no benefit to the actor. These are ignored here, largely because in such cases there is a tremendous temptation to characterize them as involving intentional harm, rather than reckless or negligent. After all, if the agent could expect nothing for himself from the act, it seems that he must have positively sought harm to others.

[xvi] *Model Penal Code*, Commentary on §2.02(2)(c), p. 237.

[xvii] Larry Alexander and Kimberly Kessler Ferzan (with Stephen Morse), *Crime and Culpability: A Theory of Criminal Law*, Cambridge: Cambridge University Press, 2009.

[xviii] Paul H. Robinson, *Causing the Conditions of One's Own Defense: A Study in the Limits of Theory in Criminal Law Doctrine*, 71 VA. L. REV. 1, 16-17 (1985)

[xix] Kimberly Kessler Ferzan, *Opaque Recklessness*, The Journal of Criminal Law and Criminology, Vol. 91, No. 3 (Spring, 2001),pp. 597-652

[xx] See *Model Penal Code* §2.06(2)(a).