2008

Telling Differences: Observational Equivalence, Externalities, and Wrongful Convictions

Manuel A. Utset
Florida State University College of Law

Follow this and additional works at: https://ir.law.fsu.edu/articles

Part of the Criminal Law Commons, Criminal Procedure Commons, and the Law and Economics Commons

TELLING DIFFERENCES: OBSERVATIONAL EQUIVALENCE AND WRONGFUL CONVICTIONS

Manuel A. Utset[∗]

We must begin with the mistake and transform it into what is true. That is, we must uncover the sources of error; otherwise hearing what is true won't help us. It cannot penetrate when something is taking its place. To convince someone of what is true, it is not enough to state it; we must find the road *from error to truth.*[1]

TABLE OF CONTENTS

1 LUDWIG WITTGENSTEIN, REMARKS ON FRAZER'S *GOLDEN BOUGH* 1e (Rush Rhees ed., A.C. Miles trans., 1979) (emphasis in the original).

I. INTRODUCTION

Society has little tolerance for two types of mistakes in criminal prosecutions: when a person who is in fact guilty is acquitted, or one who is in fact innocent is tried and convicted.[2] In both, the criminal justice system yields the wrong result, and in doing so, fails to meet its overarching goal—convicting offenders if and only if they are factually guilty of a crime.[3] What really goes wrong when a person is wrongfully convicted? One way to answer this question is to look at known cases in which innocent individuals were convicted. The evidence from reported exonerations suggests that eyewitness misidentifications, perjury, and false confessions have all played an important role in securing wrongful convictions.[4] The problem with this ex post approach is that the same procedural rules that led to these wrongful convictions have also produced the correct result in many other cases.[5] In order to get a more complete understanding of the types of evidentiary failures that produce incorrect outcomes, it is necessary to have a testable theory that makes accurate predictions of when innocent individuals are likely to be convicted. This is a relatively difficult endeavor, given numerous obstacles to identifying wrongful convictions and, thus, of gathering data about them.[6]

One possible strategy, which I take in this Article, is to focus on the sequence of observable interactions between the myriad actors brought together by a criminal act. Crimes are deeply social transactions; they cast a relational net, broad in scope and lasting in effect. When people violate the law, they create a number of explicit and implicit relationships with victims, witnesses, investigators, prosecutors, judges, juries, and innocent suspects.[7] Each of these actors will

---

[2] *See* United States v. Havens, 446 U.S. 620, 626 (1980) (stating that "finding the truth is a fundamental goal of our [criminal justice] system"); United States v. Nobles, 422 U.S. 225, 230 (1975) (identifying that two important properties of the system are assuring that "guilt shall not escape or innocence suffer") (quoting Berger v. United States, 295 U.S. 78, 88 (1935))).

[3] There are other important goals of the criminal justice system that will not directly concern us, such as protecting the autonomy or dignity of citizens. *See* 1 WAYNE R. LAFAVE ET AL., CRIMINAL PROCEDURE §1.5 (3d ed. 2007) (setting forth the various goals of criminal procedure system, including truth, providing lay participation, and respecting the dignity of citizens, and discussing various considerations when balancing them).

[4] *See* Samuel R. Gross, *Convicting the Innocent*, ANNUAL REV. L. & SOC. SCIENCE (forthcoming 2008) (manuscript at 18, on file with author) (reporting evidence for wrongful convictions in murder and rape cases and finding that eyewitness misidentification, perjury, and false confessions played important role).

[5] *See id.* at 19 (arguing that the procedures used to wrongfully convict individuals in reported exonerations "often produce accurate information").

[6] *See id.* at 20–21 (discussing the lack of data and difficulty in collecting it, and summarizing the results of existing empirical studies).

[7] Victimless crimes involve similar relationships, since they necessarily involve the community that prohibited the act in question, and, if repeated, over time can impose costs on other actors, such as family members. Additionally, some victimless crimes are only victimless to the extent that one excludes a person's own future self as a potential victim. It

undertake one or more actions between the commission of a crime and a conviction. These action sequences will intersect at different points in time, interactions that are sometimes observable, but sometimes not; who observes, and when, matters, as does the identity of those who must rely on the testimony of these observers. Eyewitnesses and crime scene investigators both interact with offenders, although at different times and through different modes of communication; and juries, in turn, interact with offenders through these informational intermediaries. Between the time that a witness observes a crime and the time that she delivers her testimony, she will have repeated occasions to communicate with investigators and prosecutors—to clarify facts and transform them into a coherent whole—a necessary exercise that may increase or decrease the accuracy of her representations to the jury. A theory of wrongful convictions should provide an account of the role played by these intertemporal interactions in securing convictions, including the types of things that can go wrong along the way.

This Article develops a theory of wrongful convictions based on the concept of behavioral or observational equivalence. As a general matter, two actors are observationally equivalent if a third party cannot distinguish between them, based on its observations of their actions.[8] The theory begins with the assumption that offenders communicate with their environment, both knowingly and unknowingly, through a series of direct and indirect observation events. In other words, the evidence set used in criminal prosecutions bottoms out in a series of events or actions, triggered by the criminal act, that are observable by one or more actors. These observations may occur concurrently, as when someone witnesses a crime, or later in time, when the authorities gather and analyze physical evidence. The state will use this evidence trace to determine whether a crime occurred and the identity of the offender. Under this approach, a wrongful conviction is the end-product of a sequence of observations that when put together leads a jury to conclude: (1) that a person committed a crime; and (2) that the innocent defendant is that person.

It is important to treat these as two distinct steps, since conflating them obscures some important causes of mistaken convictions. To model the first step, I will assume that a crime is defined by a set of hypothetical action sequences that if observed are sufficient to trigger a violation. In order to show that a crime occurred, a prosecutor has to establish that one of these hypothetical action sequences is equivalent to a real one that was performed by *someone*—who I will call the hypothetical offender. I use the concept of a hypothetical offender to model the second step in securing a conviction: a prosecutor has to show that the defendant at trial is equivalent to the hypothetical offender, based again on the defendant's observable behavior. The notion of equivalence plays an important

---

is in avoiding harm to these "victims" that the paternalistic lawmaker finds justification. *See* Manuel Utset, *Hyperbolic Criminals and Repeated Time-Inconsistent Misconduct*, 44 HOUS. L. REV. 609, 631–32 (2007).

    [8] *See infra* Part II.D.4. (developing concept of observationally equivalent agents).

role in both steps, and the principal contribution of this Article is to show that the way society defines these equivalence relations plays a critical role in convicting innocent defendants.

I develop the concept of observational equivalence in Part II, based on one used by engineers to specify and design real-time concurrent systems. I first show that the criminal justice system is an example of a concurrent system. After that, I introduce three design principles developed by engineers to predict and avoid system failures due to indeterminacy and complexity. In doing this, I define the concept of an observation event and show how it is used to determine whether or not two actors in a concurrent system are equivalent. I then use this equivalence measure to develop a model of wrongful convictions in Part III.

After doing this, I turn, in Part IV, to various legal implications of the model. Among other things, I show that the likelihood of a wrongful conviction due to a witness's cross-racial misidentification can be reduced by making use of the witness's very weakness: her inability to tell apart individuals of other races. In short, two mistakes can be better than one. The question of equivalence also plays a role in post-conviction challenges. Commentators and policymakers have assumed that all innocent individuals who are wrongfully convicted are equivalent along one dimension: they should all be exonerated and released from prison. One plausible exoneration rule is the following: if society determines with one hundred percent certainty that a prisoner is innocent, it should immediately exonerate and release her. However, using very plausible assumptions, I show that innocent offenders residing behind a veil of ignorance would reject such a rule. In other words, they would choose to keep some of these innocent individuals imprisoned. The argument makes use of the fact that wrongful convictions will be overturned sequentially, over time; which means that the order in which innocent individuals are released matters. For example, a recidivist whose most recent conviction is overturned and offends again will impose a negative externality on the innocent still in prison. The intuition is straightforward: judges and lawmakers who observe the recidivist's new crime will have a greater incentive to increase the burden faced by subsequent prisoners claiming innocence and seeking exoneration. These hurdles matter because barring clear DNA evidence, proving innocence with one hundred percent certainty is usually impossible. Part V provides some conclusions.

II. CONCURRENCY AND THE CRIMINAL JUSTICE SYSTEM

Evidence from past exonerations indicates that erroneous witness identifications, perjury, and false confessions play an important role in wrongful convictions.[9] These are all material errors, in that a single instance can lead a jury to convict an innocent person. While our understanding of these errors is not in any way complete, it is much further advanced than the type of error that I am concerned with in this Article: sequences of smaller, nonmaterial errors that—in the aggregate—can lead witnesses, investigators, prosecutors, judges, and jurors

---

[9] *See* Gross, *supra* note 4, at 18.

down a path, sufficiently divergent from the truth, that it culminates in a mistaken conviction. This latter type of sequential evidentiary mistake has not received systematic attention from scholars and policymakers. One reason is that these smaller errors are difficult to observe, ex ante and ex post, and thus harder to verify empirically;[10] another is that they do not afford easy or obvious solutions.

My general argument is that a comprehensive theory of wrongful convictions has to provide an account of the way in which a sequence of nonmaterial errors can lead to a wrongful conviction. The theory that I develop in this Article provides such an account, one based on the concurrent nature of the criminal justice process. I begin this Part by describing the characteristics of the criminal justice system that contribute to this problem. I then turn to the general problem of how an ideal lawmaker would go about specifying and implementing a set of complex legal rules. In the last two subparts, I set forth a simple model of the criminal justice process and the three design principles used by engineers to predict and detect cumulative errors in real-time, concurrent systems.

A. The Concurrent Nature of the Criminal Justice System

The criminal justice process begins when a lawmaker enacts a legal rule and ends when a person convicted under that rule exhausts all avenues for overturning the conviction; along the way, the actions of numerous individuals will impact the process. Parties brought within an offender's relational web will sometimes act alone, beyond the gaze of others, and sometimes jointly, either with cooperative or strategic intent. These properties of the criminal justice system make it a concurrent system, one composed of actors or components that are sometimes required to act in unison or concurrently, but also have the ability to act independently. Concurrency problems and errors can arise whenever the independent actions of one or more actors within a system are not fully observable by others.

I begin this subpart by describing the relationship between legal errors, on the one hand, and the complexity and indeterminacy of the criminal justice process on the other. I argue that these errors are not unique to the justice system: they are common in any real-time concurrent system, and thus have received close attention from engineers. I then describe certain important features of the relationship between discretion, indeterminacy, and observational gaps. I finish the subpart by providing a stylized account of how the sequential and concurrent interactions of actors can lead to cumulative errors and wrongful convictions.

---

[10] Of course, just because they are difficult to identify does not mean that they do not play a role in wrongful convictions, particularly for nonmajor crimes. *See* Gross, *supra* note 4, at 22 (arguing that while there is little evidence of wrongful conviction for lesser crimes, they are likely to be due to smaller, administrative and investigative errors).

1. Indeterminacy and Complexity

How can society check the reliability of the criminal justice system? As a general matter, lawmakers cannot anticipate every possible context in which a rule may, or should, apply, or draft legal rules that are guaranteed to be interpreted in the manner intended. A legal rule is indeterminate if applying it to the same set of facts leads to different results. An important source of indeterminacy in the criminal justice system is the large amount of discretion afforded to state actors. Discretion is a difficult concept to define precisely, but at a minimum involves a person's ability or freedom to choose between two or more possible actions. For example, prosecutors may exercise discretion when deciding whether or not to charge a suspect, and judges, when sentencing defendants, although, in both cases, their discretion is subject to a number of procedural and legal constraints.

The problems created by the inherent indeterminacy of legal rules are exacerbated by the overall complexity of the criminal justice process.[11] This complexity is due to the large number of actors involved and the complicated, nontransparent nature of their relationships.[12] The greater a system's complexity, the greater is both the potential for mistakes and the general difficulty of predicting how substantive and procedural rules will operate. Errors due to indeterminacy and complexity are common whenever concurrent systems operate in real time. Engineers routinely design real-time systems that like the criminal justice system have a low tolerance for error, such as hospital equipment, air traffic control systems, and nuclear power plants. These systems are composed of semi-autonomous nonhuman actors, and the myriad human beings that interact with them. Given the intertemporal, dynamic nature of these interactions, a designer has to have a good sense of how each actor's independent, nonobservable behavior affects their relationship with others in the system. As we will see below, engineers have developed numerous methods to test and simulate the behavior of concurrent systems to identify and weed out potential errors, and reduce their indeterminacy.[13]

---

[11] *See* HERBERT A. SIMON, THE SCIENCES OF THE ARTIFICIAL 183–84 (3d ed. 1996) (defining complex systems as "made up of a large number of parts that have many interactions").

[12] *Id.* at 184 (arguing that a system's complexity will increase whenever, given "the properties of the parts and the laws of their interaction, it is not a trivial matter to infer the properties of the whole").

[13] These formal methods were developed precisely to deal with the problem of designing systems in which the joint activities of actors mattered and can lead to implementation errors and indeterminacy. Importantly, the actors modeled include both artificial computer processes and the human beings who interact with them. *See* C.A.R. HOARE, COMMUNICATING SEQUENTIAL PROCESSES 72–79, 108–112 (2004) (setting forth formal methods used to specify, implement, and test concurrent systems to deal with indeterminacy and system deadlocks); ROBIN MILNER, COMMUNICATION AND CONCURRENCY 195–203 (1989) (introducing the calculus of communicating systems, which models interactions between agents as a series of synchronization activities); STEVE

2. *Observations and Discretion*

To understand the relationship between discretion and indeterminacy in the justice process, it is useful to begin with a definition of determinism. An actor is deterministic if whenever it is in a position to choose between two or more actions that choice is made by the environment. This means that her choice is either made or indirectly controlled by another actor who has the ability to observe her behavior (the environment).[14] Observation plays a role in reducing an actor's autonomy whenever: (1) the actor has the ability to choose freely, when acting anonymously; and (2) the observer has the power to limit or completely extinguish the number of choices available to her. For example, the level of discretion of an offender, the police, and prosecutors will depend on whether their behavior is observable, as well as the identity and power of their observers. When the police observe offenders in real time, they will reduce the set of choices available to them, as will observations by witnesses, although to a lesser extent.

It is of course well-known that observation can limit discretion;[15] however, legal commentators have not given any real attention to the role played by sequences of observations, made by different actors, at different times, and their overall impact on the criminal justice process; not just on the discretion of parties, but also on the reliability of the system. Since defendants are convicted or acquitted based on these sequences of observations, they are ultimately the source of wrongful convictions. In other words, the path from the commission of a crime to a conviction is defined by a sequence of actions by myriad parties, where observing another person is one of these actions. My argument is that in order to understand the path from a crime to a wrongful conviction, we need to have a full understanding of these observation events. Moreover, by isolating one type of action—observation events—we in turn reduce the overall complexity of the system, making it easier to identify potential sources of errors.

Observability gaps lead to indeterminacy, complexity, and mistaken results in other types of concurrent systems. As a result, engineers designing these systems have drawn a clear distinction between observable and nonobservable behavior. Legal scholars and lawmakers can learn from the experience of engineers and import some of their tested methodologies to specify, design, and implement more reliable substantive and procedural rules[16]—rules that are better calibrated to the

---

SCHNEIDER, CONCURRENT AND REAL-TIME SYSTEMS: THE CSP APPROACH (2000) (discussing theory of real-time concurrent systems).

[14] *See* HOARE, *supra* note 13, at 81 (defining a deterministic process as one whose behavior is completely determined by the environment, or one in which each action is observed by another).

[15] In Justice Brandeis's oft-quoted terms: "Sunlight is said to be the best of disinfectants; electric light the most efficient policeman." LOUIS BRANDEIS, OTHER PEOPLE'S MONEY 92 (1914).

[16] A specification is a description of the desired properties or behavior of a system being built. A design is a blueprint or abstract plan of a system meeting the specification. Finally, an implementation is the actual instantiation of the system. There is often more

particular problems of concurrent systems. Beginning in the next section, I develop an engineering-based model of the criminal justice process; but it is first helpful to provide an account of the relationship between observation chains, cumulative errors, and wrongful convictions.

3. Cumulative Mistakes and the Path to Wrongful Convictions

Before the authorities can mount a criminal prosecution, they have to determine: (1) that a crime occurred and (2) the identity of the offender. Knowing this, a rational offender will attempt to limit and disguise—in short, manage—the evidence trace that she leaves behind. This evidence trace is a record of the offender's interaction with her environment, with the people and objects that she meets along the way. These third-party observers witness the offender's behavior, record it, and eventually transfer the information to investigators, who combine this testimony with their own observations of the crime scene and of potential suspects. Once they are finished, investigators will package the evidence and pass it along to a prosecutor, who will put it together with her own observations and present the amalgam to a jury, or at least those portions that a judge admits into evidence.

This process will culminate with either a correct or mistaken conviction (or no conviction at all). To understand wrongful convictions it is thus necessary to know what can go wrong, the timing and types of errors, and the procedures used to detect and correct them. Errors will undoubtedly creep in along the way, as these actors observe, interpret, and record the offender's evidence trace, identifying logical and evidentiary gaps and closing them by drawing inferences. This gap-filling exercise is similar to that employed by courts when they supply terms missing from a contract,[17] although in criminal cases, the story presented at trial has many more authors and points of view.

In summary, large errors such as eyewitness misidentifications or perjury are important sources of wrongful convictions, but so are sequences of relatively small, undetected errors that, in the aggregate, allow prosecutors to convince a jury to convict an innocent defendant. Large errors are easier to identify and avoid; a series of small ones, are far more insidious, and can lead to much greater social harm. Not only are they more difficult to detect before an innocent defendant is convicted, but also afterwards. To convince a judge to reopen the case, an innocent

than one way to implement a system; but the goal is to achieve one that can be shown to meet the specification, and thus behaves correctly. One may then choose between different equivalent implementations using other metrics, such as cost-effectiveness, robustness, or amenability to future modifications. *See, e.g.*, Carlo Ghezzi et al., *Software Qualities and Principles*, *in* THE COMPUTER SCIENCE AND ENGINEERING HANDBOOK 2278, 2281–82 (Allen B. Tucker ed., 1997) (describing the often iterative process from specification to implementation).

[17] *See* Manuel A. Utset, *A Theory of Self-Control Problems and Incomplete Contracting: The Case of Shareholder Contracts*, 2003 UTAH L. REV. 1329, 1381–82 (discussing contract gap-filling strategies used by courts).

convict will have to identify these errors and provide an account of the way in which they coalesced to materially distort the evidence set. A convict who has discovered a few of these mistakes may need an evidentiary hearing to identify the rest; but a judge may not grant one until the group of errors, known so far, are sufficiently material.

B. *Specifying and Implementing Legal Rules*

Specifications are the first line of defense when designing complex concurrent systems, whether of laws, or other types of components; they help identify contexts in which the application of a rule may lead to indeterminate or mistaken results, and allow designers to check whether the implementations they settle on are reliable and robust. In order to design a reliable criminal justice system, an ideal, omniscient lawmaker would first specify how the myriad actors involved in the criminal justice process should behave, or at the very least determine, for each one, the actions that they are prohibited or required to take.

The lawmaker would do this not just for substantive rules, but also procedural ones. Procedural rules tell state actors how they can or cannot behave when they are involved in the criminal justice process. These actors will produce both intermediate and terminal outcomes, such as detecting and successfully prosecuting an offender, securing a confession, or exonerating a person who was mistakenly convicted. In designing procedural rules, the lawmaker will need to determine what would count as correct or incorrect terminal outcomes—for example, the acquittal or conviction of an innocent person. While the final result is of obvious import, a lawmaker also needs to have a good sense of what should count as correct or mistaken intermediate results, and of the cumulative results produced by different paths of behavior.

Once the lawmaker has created a specification of desired actions and outcomes, it will turn to the implementation phase. Its goal is to draft substantive and procedural rules that meet those specifications.[18] It is at this juncture that the complexity of the system and indeterminacy of legal rules can create problems. To deal with these, the lawmaker can try to simulate or predict how the legal rule would operate under various scenarios, test the system on an ongoing basis, and adopt legal rules that are easy to modify if these tests indicate that they are unreliable.

---

[18] A designer's ability to prove a system's correctness—that it is guaranteed to meet the specifications—can be important where system errors can produce significant harm; however, if the system is at all complex, it is virtually impossible to prove that the whole system will yield correct results. As a general matter, the best that system implementers can hope for is proving that certain intermediate results will meet the specifications, and that the overall system is reliable in that its behavior stays within a specified interval of acceptable error. For example, computer scientists have developed a number of formal methods to prove that *parts* of their software systems are guaranteed to yield the correct result. *See* Ghezzi et. al., *supra* note 16, at 2282.

C. *Identifying the Relevant Actors and Actions*

The complexity of the criminal justice process is due to the large number of actors involved and the way in which they interact, including the number and timing of these interactions, and the general inability of parties to fully observe each other's behavior. As a first step in reducing this complexity, a designer can create a typology of actors and actions available to each. I will assume that the criminal process begins when an offender commits a crime, and that in addition to the offender and victim, it may involve one or more witnesses, investigators, prosecutors, judges, and jurors. We can begin to model these actors by assuming that they will each have a set of actions available to them. For example, an offender may or may not violate the law, prepare for the crime ahead of time, or engage in behavior to avoid detection. A witness may observe a criminal act, report it to the authorities, and testify in trial. Investigators may take affirmative steps to detect crimes and offenders, and identify victims and witnesses. We can do the same with others involved in the process, including, as we will see later, innocent suspects brought into a crime's relational web.

The content of these action sets will vary depending on the particular contexts and aspects of the problem that one is interested in modeling. I am interested in the sequence of actions that culminate in the construction of the evidence set that is used to convict or acquit a defendant. At any one point in time, each person will have a history, setting forth a set of actions and the order in which they occurred; in between any two time periods, a person's history is a record of each action that the person took during that period. Because the order of these actions or events matters,[19] I will refer to them as *action sequences*.[20] It is often possible to create a slice of an actor's complete history using one or more properties to isolate an action sequence of interest. Importantly, not all of this history is observable by other actors, and a person's statements about it are not always credible.[21] As we

---

[19] *See* Schneider, *supra* note 13, at 85 (stating that, in concurrent systems, an important property of an agent's behavior is the "occurrence of events in the right order, and that events do not occur at inappropriate points").

[20] A sequence of *n* objects or components is an ordered set (or *n*-tuple), where the components are identified by their position in the ordered set. More formally, if $a_1, a_2, \ldots, a_n$ are the components, then a sequence is an ordered *n*-tuple, $(a_1, a_2, \ldots, a_n)$, where *n* is the length of that sequence. For example, the following two sequences are not equivalent: *A = (a, b, c, c)* and *B = (b, a, c, c)*. Moreover, unlike sets, where we are only concerned with who is a member, in a sequence we not only care about the identity and order of the components, but also about how many times they occur. In other words, under set notation *A* and *B* would be the same set: *{a, b, c}*. *See, e.g.*, HARRY R. LEWIS & CHRISTOS H. PAPADIMITRIOU, ELEMENTS OF THE THEORY OF COMPUTATION 10 (2d ed. 1998) (defining sequences).

[21] More formally, an agent's complete action set is composed of the observable action set and one or more internal actions available to it that cannot be observed by the environment. It is sometimes possible for other agents to infer the content of internal events by observing an agent's subsequent external behavior, but it is these inferences that lead to errors. *See* ROBIN MILNER, COMMUNICATING AND MOBILE SYSTEMS: THE π-CALCULUS 38,

will see, these *observability gaps* are an important source of errors in the criminal justice process. For example, when an offender is arrested, the authorities will gather crime-related evidence to determine the offender's history leading up to, during, and after the crime; but this evidence is limited to those portions of the offender's behavior that are observable. In order to understand how observability gaps can lead to cumulative errors in the criminal law process, we must get a better grasp of the role played by observation events in the design of concurrent systems.

D. The Role of Observability Gaps

If a lawmaker were to hire an engineer to specify, design, and implement a criminal justice system, how would the engineer approach the task? What design methodologies would it use to predict and detect cumulative errors? In this subpart, I describe the design principles used by engineers to model concurrent systems. Their general approach for reducing indeterminacy and observability gaps is simple: make all material system actions joint observation events; if an action is not observable, it is ignored—that is, the designer assumes that these actions have no information content that is useful for predicting the system's behavior.

1. Three Design Principles for Concurrent Systems

Observability gaps lead to indeterminacy and mistaken results in other types of concurrent systems; for example, nonobservability can lead to deadlocks and coordination failures.[22] The key design breakthrough by engineers was the realization that in many cases they could simply ignore nonobservable behavior, as long as they did so in a principled, well-thought-out manner. This proviso led them to adopt three general design principles. First, they treat each actor in a system as a "black box"—self-contained entities whose behavior is either internal or external.[23] When two actors interact, they do so via their external, observable behavior; their internal behavior remains hidden. Second, engineers adopt the categorical assumption that actors are completely defined by the set of observation events (observable actions set) in which they can participate, or equivalently, by the manner that they can interact with their environment.[24] The third assumption is that

---

52–54 (1999) (describing action set of external and internal actions and proposing mechanism that allows one to disregard some occurrences of internal events).

   [22] *See, e.g.*, HOARE, *supra* note 13, at 59–60 (describing deadlocks that can occur when two agents must share the same resources).

   [23] *See* MILNER, *supra* note 21, at 13 (1999) (describing agents as "black boxes" that can be distinguished only by observing their external behavior).

   [24] An agent's environment is a catch-all term for the set of all agents in a system that can in some way observe the first agent's behavior. *See* A.W. ROSCOE, THE THEORY AND PRACTICE OF CONCURRENCY 13 (1998) (stating that an agent is "completely described by the way it can communicate with its external environment"); SCHNEIDER, *supra* note 13, at 1 (describing process algebra approach as one where components are modeled as

two actors can observe each other only if they participate in a joint activity or interaction; and the act of participating will always produce some observation or communication.[25] These observations are not necessarily complete or fully reliable, and it is possible for each participant to observe different things.

In order for an interaction or joint event to occur, agents have to be able to communicate or otherwise synchronize their behavior. But these communications are not limited to transfers of information, signals, or messages, as those terms are used in economics and information theory.[26] Instead they bottom out in some sort of joint *observation event*, one in which the information transferred can take many forms and may even disappear after the event occurs.[27] Under this approach, a victim "observes" an offender's actions by experiencing the harm; in other words, the offender and victim are involved, by necessity, in a joint observation event, although harming another is not usually considered a form of communication.[28]

2. Observations Events

When one combines these three design principles, the picture of concurrent systems that emerges is one in which the observation event is what gives a system its structure and stability: every relevant action is an interaction, and every interaction involves some form of observation.[29] This means that two agents can engage in a joint activity only if it is part of both of their observable action sets.[30] This sort of joint observation event is often described as a "handshake" synchronization,[31] where the latter term helps underline the importance of

"independent self-contained entities with particular interfaces through which they interact with their environment").

[25] In other words, a necessary condition for person *A* to observe an action of person *B* is that she interacts with *B* in some specified manner; and through this interaction *B* in turn observes *A*. *See* MILNER, *supra* note 21, at 28 (stating that an agent observes the action of another agent by interacting with it).

[26] *See* DAVID M. KREPS, A COURSE IN MICROECONOMIC THEORY 629 (1990).

[27] *Cf.* ROSCOE, *supra* note 24, at 8–9 (1998) (stating that communications should be thought of as a "transaction or synchronization between two or more processes rather than as necessarily being the transmission of data one way").

[28] Note that it is possible for an offender to take an action at time *t*, but no one notices it until time *t + 1*. While the crime was technically committed a time *t*, the machinery of the state will remain dormant until the observation at *t + 1*. Statutes of limitations are sometimes tolled for exactly this reason; an offender who has concealed her crime in particular ways may find that the state does not consider that the crime "occurred" until it is discovered.

[29] *See* MILNER, *supra* note 13, at 12 (stating that under this approach "the behavior of a system is exactly what is observable, and to observe a system is exactly to communicate with it").

[30] *See* HOARE, *supra* note 13, at 46 (two agents can engage in a concurrent event only if it appears in both of their observable action sets).

[31] Examples of synchronization events of this sort include "passing of a baton in a relay race; the delivery of a registered letter; the closure of a contract; becoming married,"

simultaneity. In other words, a joint event can occur only when all participants have reached a point in their action sequences in which participation is possible; at which time they will either follow through or choose an alternate action. This means that even if all parties are at a point in which they can participate, one or more of them may have the discretion to choose another action or delay executing the joint event; this can lead to divergences or deadlocks in the system, and are thus an important source of error.[32]

3. Two Sources of Indeterminacy

Observability gaps allow indeterminacy to creep in along two dimensions. First, if two actors are unable to observe each other's independent behavior, their ability to engage in joint activities can be negatively affected. Nonobservable behavior can make it difficult or impossible for them to predict whether they will reach a point in their action sequences in which they can act jointly, or ascertain if they have already reached that point. Second, synchronization failures can occur whenever an actor is able to engage in a required joint event but can refuse to do so. These two types of coordination failures are what lead to deadlocks, divergences, and other system errors.

As a result, a designer has to have a way of predicting how the behavior of each actor will evolve during the time period in which they will have to engage in one or more joint events. To do this, a designer will specify the set of observable actions that each actor is able to take, and the order in which those actions may occur. By limiting themselves to these observable action sequences, engineers commit themselves to a design strategy in which the only actions relevant to two actors are observable ones. To bridge the gap between observable and nonobservable behavior, engineers have developed a number of techniques that make important use of the concept of equivalence. While they have developed various equivalence tests, I will focus on a well-established approach, one that defines equivalence using the concept of traces of observable behavior: observational equivalence.

---

and more generally, any action involving two or more actors whose behavior intersects temporally in a meaningful way. *See* SCHNEIDER, *supra* note 13, at 31. *See also* MILNER, *supra* note 21, at 36 (describing handshake synchronizations as atomic or indivisible actions of two or more agents, where data may or may not be transferred and in which there is no necessary directionality in the "communication").

[32] *See* HOARE, *supra* note 13, at 49–50, 94 (defining "deadlocks" as occasions in which two components are required to undertake a joint event, but one or both are unable to do so; and "divergences" as situations in which an actor who has the ability to engage in a joint event also has the concurrent ability to undertake an infinite sequence of other actions without committing to that joint event).

4. Observational Equivalence

As a general matter, two individuals are equivalent if a third party cannot tell them apart.[33] But this statement underscores the general problem of dealing with questions of equivalence: there is no a priori or self-evident way to tell things apart, at least not one that is applicable in all contexts.[34] As a result, the first step in any equivalence theory is to identify the set of agents to be compared and the reason for comparing them, since this will influence the dimensions along which they will be ideally equated or distinguished.[35]

Under the observational approach, a designer has to determine when it can safely conclude that two possible instantiations of an actor *A* and *A´* are equivalent. To see how the designer would approach this problem, suppose that, in between any two observable actions, both *A* and *A´* are each able to undertake one or more actions that are not observable by their environment. If each one has a different set of nonobservable actions, then a designer will need to determine whether the two actors are sufficiently equivalent and thus interchangeable. And here is the catch: it must do so based only on the observable behavior of *A* and *A´*. As we will see in Part III, a designer and a jury face the same task: to take a record of observations, knowing that it does not include the complete history of behavior—that is, it does not include nonobservable actions—and determine whether they can, with confidence, conclude that two actors are the same; in the case of the jury, that the defendant is equivalent to the actual offender who committed the crime.

Since only observable actions are relevant, under the concurrency model, an actor is fully defined by three sets of *observable* action sequences. The first comprises the observable action sequences that are available to her, each of which is represented by a trace. Under the model, the *trace* of an action sequence is a record of an actor's observable behavior, in the order in which those actions occur, and the set of traces is the collection of all action sequences that may possibly be

---

[33] Philosophers have dealt with this behavioral equivalence problem under the rubric of the "identity of indiscernibles." *See, e.g.*, Gordon Nagel, *The Identity of Indiscernibles*, 73 J. PHIL. 45, 45–50 (1976). More generally, philosophers have long wrestled with the question of what it means for a person to be the same person over time. The general problem is one of identifying those characteristics or factors that if they remain unchanged over time would allow one to conclude that agent *N* at time *t* is the same agent *N* at time *t + 1* and not some other agent *N´*. *See* ROBERT NOZICK, PHILOSOPHICAL EXPLANATIONS 29–37 (1981).

[34] *See* ROBERT NOZICK, INVARIANCES: THE STRUCTURE OF THE OBJECTIVE WORLD 79–80 (2001) (discussing general problem of choosing invariance conditions to deal with the issue of objectivity).

[35] An equivalence relation on a set is an operation that divides (or partitions) that set into mutually exclusive subsets in which the members are equivalent in some way. More formally, an equivalence relation "~" is a binary relation on a set *S* which is reflexive (if *a* is a member of *S*, then *a* ~ *a*), symmetric (if *a* and *b* are members of *S*, then *a* ~ *b* and *b* ~ *a*), and transitive (if *a*, *b*, and *c* are members of *S*, and *a* ~ *b* and *b* ~ *c*, then *a* ~ *c*). *See* JOHN R. DURBIN, MODERN ALGEBRA 51–52 (1979).

recorded.[36] Some actors may have the ability to undertake an infinite number of internal nonobservable actions and never follow through with an interaction. These divergent action sequences are a second distinguishing characteristic and are referred to as an actor's set of *divergences*. The third defining characteristic is an actor's discretion to refuse to participate in joint activities; these refusals will occur in real time—when others are waiting for that actor's participation. The set of actions that an actor is able to refuse is that actor's *refusal set*.

Under this model, two actors are observationally equivalent if and only if they have the same trace, divergence, and refusal sets.[37] We can summarize the general approach as one that starts with the broadest possible equivalence class—trace equivalence—and then applies two transformations to the set of traces. In other words, two trace-equivalent actors may or may not be equivalent along the divergence and refusals dimensions. The goal of these transformations is to weed out actors who may have the same traces but are sufficiently different along the other dimensions.

5. Conclusion

While the concurrency approach may appear strange at first, the rationale is clear. Concurrency problems are a product of the independent nonobservable actions of individuals; moreover, an observation can occur only if there are two or more actors—the observer and observed. This means that all observations are joint events, and the easiest way to build them into a model that captures the overall behavior of a system is to require that, whenever two actors need to share, disclose, signal, or in any way communicate or transfer information, they do so through specified interactions or synchronizations.

There are two principal benefits of the concurrency model. First, it provides a set of tools that greatly simplify the task of describing and reasoning about complex interactions within intertemporal contexts. Many of these systems are in constant operation for very long periods of time, and are required to provide real-time solutions when people interact with them. Modeling them using joint

---

[36] *See* SCHNEIDER, *supra* note13, at 85–86 (describing traces as records of the events that may occur in connection with an agent's action sequence, as if some omniscient observer were "recording all events as they are observed").

[37] More formally, suppose that at time *t*, actor *A* is in context *C* in which a joint action *x* is available—that is, there is a second actor B who is in a position to engage in that joint action. There are three possible results. First, *A* is at a point in its action sequence in which it can undertake action *x*, and it has no choice but to follow through with it. Second, *A* is at a point in which it can do *x*, but it has the discretion to refuse to engage in *x*. Third, *A* is at a point in which it can do *x*, but it also has the ability to undertake an infinite number of internal actions before it has to follow through with *x*. The first scenario is a deterministic one and is captured by *A*'s trace set; the second, involves *A*'s refusal set; and the third is an example of a divergence. It is the last two that create the potential for nondeterministic results and system errors—for example, *B* may not be able to proceed with other actions until *A* agrees to do *x*. *See* Schneider, *supra* note 13, at 171–76.

observation events helps reduce complexity by indicating the precise points at which the actions of two or more autonomous agents are required to intersect.[38] Second, the model makes it easier to specify, design, and implement systems, and to determine places in which unreliable interactions may occur. By treating all observable behavior in the same manner, it is much easier to compare the expected behavior of systems, given that what is important is not the number of components that are interacting within the system but the observable result that they produce as a group.[39] We will now use the concurrency design principles introduced in this Part to develop a model of wrongful convictions based on the concept of observational equivalence.

III. OBSERVATIONAL EQUIVALENCE AND WRONGFUL CONVICTIONS

This Part develops a simple model of wrongful convictions; the general intuition behind it is the following. A crime is defined by a set of action sequences, any of which triggers a violation of the law; I refer to them as sequences because the order in which those actions occur is often relevant. An *offender* is a person who executes a sequence of actions that is equivalent to one of those illegal action sequences. Everybody else is an innocent person, any of whom is susceptible to a wrongful prosecution and conviction, although with varying levels of probability. The probability of a wrongful conviction is a function of the observable action sequences of an innocent defendant before, during, and after the actual offender committed the crime; and the probability of a correct conviction is a function of the offender's observable action sequence during the same period. More specifically, in order to convict a person correctly or incorrectly, a prosecutor will present at trial an evidence set, representing an action sequence that convinces the jury, beyond a reasonable doubt, that: (1) a person committed a crime; and (2) that that person is the defendant. In short, the prosecutor will need to convince the jury that the defendant undertook a sequence of actions that is equivalent to one of the action sequences that triggers a violation of the law.

My model attempts to isolate those parts of the process that can lead to a wrongful conviction, doing so along three dimensions. The first involves the

---

[38] Although simultaneity will always be a factor in a synchronization event, joint activities can also involve other relevant dimensions, including space and the identity of the agents. In fact, in many instances it is unnecessary to know the precise time at which a joint event occurs, just that it occurred within some time interval. It is simultaneity—their intersection at some point—which matters. *See* ROSCOE, *supra* note 24, at 6 (arguing that handshake synchronizations make "precise timing of less concern, since if one end of the communication is ready before the other, it will wait").

[39] As Professor Tony Hoare, one of the pioneers in the area, argued: "it is best to forget the distinction between processes, environments, and systems; they are all of them just processes whose behavior may be prescribed, described, recorded and analyzed in a simple homogeneous fashion." HOARE, *supra* note 13, at 45. *See also* SCHNEIDER, *supra* note 13, at 91 (describing role of compositional design in achieving semantic uniformity and reducing complexity).

determination of when an action sequence is sufficient to trigger a violation of the law. The second dimension involves the set of problems that can arise during the process of aggregating observations into the evidence set introduced at trial. The third dimension involves the process through which a jury is led to believe that an innocent person violated the law; in short; that leads them to conclude that an innocent person and the offender are the same person.

A. The Crime-Equivalence Class

Legal rules create a set of instructions for future users, telling them how they need to behave in one or more contexts. These instructions act as a set of specifications for the criminal justice system, describing what counts as behavior that violates the law, and setting forth the procedures that the authorities must follow to achieve legally valid convictions. In this subpart, I argue that a crime is fully specified or defined by a set of action sequences, any of which is sufficient to trigger a violation of the law. I refer to this set as the *crime-equivalence class* for reasons that will subsequently become more obvious.

A person violates a criminal law by taking a prohibited action (or failing to take a required action),[40] where such behavior is accompanied by a specified state of mind, such as intent, knowledge, recklessness, or negligence.[41] These two events—the act and the state of mind—must occur concurrently.[42] Moreover, the external physical action and internal mental event will in most instances involve a series of subactions; these must occur in the right order and together will constitute an action sequence that describes the occurrence of a crime event. In other words, if a hypothetical jury were to fully observe that person's behavior in real time, it would interpret it as the type that, when evaluated using the instructions provided

---

[40] This action requirement is usually referred to as the actus reus of a crime. While this Article will refer primarily to prohibited actions, the same general conclusions apply to contexts in which a person fails to take a required action.

[41] The state of mind or mens rea requirement is a type of activity or conduct that acts as a sort of proxy for the blameworthiness of the person's behavior. *See* MODEL PENAL CODE § 1.13(5) (1962) (defining conduct that triggers legal responsibility as involving both an act *and* a state of mind). Additionally, some crimes require that the act (or omission) produce a certain result or meet other conditions, such as the existence or absence of special characteristics or circumstances surrounding the perpetrator's conduct.

[42] There will be no criminal liability when (1) the state-of-mind event occurs first, comes to an end, and then the offender acts; or (2) the act occurs and comes to an end, before the state-of-mind event. The difficult question in these cases is determining when the two events are sufficiently concurrent to make the offender responsible. *See* JOSHUA DRESSLER, UNDERSTANDING CRIMINAL LAW 197–99 (3d ed. 2001) (discussing problem of temporal concurrency in mens rea and actus reus). Note that here we are treating the offender as a system made of two components: a body that engages in observable behavior and a mind whose actions are not observable. *See* NOZICK, *supra* note 33, at 458–59 (discussing general mind-body problem).

by the judge, proves beyond a reasonable doubt that that person violated the law.[43] Although the actus reas and mens rea are the two principal actions, and their concurrence the main temporal component, it is the complete action sequence that ultimately matters. Introducing evidence of these other actions is usually needed when establishing justifications and excuses, or when showing the existence or absence of a requisite state of mind, or of attendant circumstances.[44] Finally, the order in which these other actions occur can determine whether legal liability will issue.[45]

A crime will therefore have associated with it a set of one or more action sequences, any one of which is sufficient to trigger a legal violation;[46] moreover, all of these action sequences are equivalent to each other along that dimension and thus form an equivalence class.[47] A person violates the law whenever her behavior involves an action sequence that is equivalent to one that is a member of a crime-equivalence class.[48] As we will see in this Part IV, the crime equivalence class can change over time, and may expand or contract after a person has committed an "alleged" crime, as when a judge adopts an expansive or restrictive interpretation of the types of behavior that fall under that crime. These sorts of ex post

---

[43] More generally, we can say that the *meaning* of this sequence of actions as interpreted by the jury, using the legal language provided by the judge in her instructions, is that "person *M* committed a crime *C*." This phrase can be inserted in a number of other expressions and taken to be equivalent to the more specific description of the offender's sequence of actions. For a further discussion of deriving a meaning in this manner, see RONALD FAGIN ET AL., REASONING ABOUT KNOWLEDGE 16–19 (1995) (discussing Kripke structures).

[44] *See* 1 WAYNE R. LAFAVE, SUBSTANTIVE CRIMINAL LAW §§ 5.1, 9.1 (2d ed. 2003).

[45] For example, a person who shoots someone in self-defense before she becomes aware of the attendant danger is treated differently from one who first observes and then shoots.

[46] It may also include corollary ones, necessary for establishing a justification or excuse.

[47] An equivalence class [*a*] (relative to the equivalence relation ~ and set *S*) is the subset of *S* whose elements *b* are those that are equivalent to *a* (*a* ~ *b*). Finally, "if ~ is an equivalence relation on a set *S*, then the set of equivalence classes of ~ forms a partition of *S*, where a partition of a set are the collection of subsets of *S* whose union yield *S* and which are mutually exclusive (that is, if subsets *A* and *B* are in the collection of subsets and *A* ≠ *B*, then the intersection of *A* and *B* is the empty set). *See* DURBIN, *supra* note 35, at 51–53. The members of that equivalence class may be further refined by positing that some action sequences may be different from others along other dimensions. A refinement of an equivalence class is an operation on it that separates a subset of that class in accordance to some property.

[48] A valid crime must contain at least one action sequence in the crime-equivalence class. A law that does not sufficiently specify the type of behavior that it makes illegal will fail for vagueness. *See* LAFAVE, *supra* note 44, § 2.3(a).

transformations of the content of the equivalence set create a number of potential justice concerns.[49]

It is now possible to define correct and mistaken convictions. Under the crime-equivalence model, a person is *correctly convicted* of crime *C* only if she undertook a sequence of actions that is equivalent to an action sequence that, at the time of the alleged crime, was a member of that crime's equivalence class. On the other hand, a person *M* is *mistakenly convicted* of crime *C* if: (1) her action sequence was not a member of the crime equivalence class, at the time of the alleged act (a law-specific mistake); or (2) someone undertook an action sequence which is equivalent to one in the equivalence class, but that person was someone other than *M* (a defendant-specific mistake).[50]

Finally, this approach to specifying legal rules is open to the potential criticism that it is too general, thereby ignoring a number of subtleties, as well as the particularities or nitty-gritty of real cases. But generality is in fact my aim: identifying a set of techniques that a lawmaker, judge, or lawyer can apply uniformly at different points in the process to test the overall reliability of procedural rules, or prove innocence or guilt. Additionally, I want to identify methods that are general enough to apply to other areas in the law that have complex collections of legal rules, such as those found in securities, tax, environmental, and administrative law.

B. Composing Evidence Sets

In this subpart, I analyze the process by which sequences of observation events are transformed into evidence sets. I will adopt the assumption that a lawmaker is trying to solve the following design problem: creating a set of rules for putting together an *evidence set* that, when introduced at trial, will lead to a conviction whenever the defendant committed the crime for which she is being tried, and an acquittal in all other cases. In particular, I argue that evidence sets contain one or more sequences of observations and that whether cumulative errors emerge and lead to mistaken convictions will depend on: (1) the order in which observation events occur, (2) the identities of the actors involved in observation events, and (3) the manner in which the content of these observations are interpreted, recorded, and transferred.

---

[49] Constitutional limitations on ex post facto laws and publicity requirements for crimes are two ways of limiting the number of potential ex post transformations of the crime equivalence class. *See infra* Part IV.D.

[50] A "wrongful" conviction of a factually guilty offender is not a mistaken conviction under this definition, since such a conviction is really a kind of second-order "mistake": a failure to convict using the correct procedures.

1. The Evidentiary Process

Once investigators or prosecutors have identified the equivalence class for a crime, they will need to gather evidence showing that a crime occurred and the identity of the offender. In order to show these two things, the prosecutor will have to introduce an evidence set composed of one or more observation events. In other words, the only real way to show the existence of a crime and tell apart the guilty from the innocent is by introducing testimony of those who experienced the criminal act directly, by observing the behavior of others—for example, the offender and victim—or indirectly, through the physical evidence left behind at the scene.[51] An initial observer will transmit what she saw to the police, an interaction that is itself an observation event, as is that between the police and prosecutors, and, eventually, prosecutors and witnesses with juries. It is along this chain of observation events where distortions, misperceptions, and lies will lodge themselves.

To avoid cumulative evidentiary errors, one needs to understand how they arise and know how to identify them in real time to correct them before they gather force; the longer errors persist and the greater the number of actors who accept them as the truth, the more difficult they become to dislodge. And as they take root, they will create reverberations down the evidence chain, as parties attempt to mold new, but conflicting evidence into a coherent whole.[52] By the time that an innocent prisoner challenges her conviction, she will face a formidable evidentiary hurdle, due not just to the persistence and wide acceptance of untruths, but to a chain of certifications by the whole apparatus of the state—the police, prosecutor, judge, and jury.

2. Building Cumulative Traces

Given that under the concurrency approach, the only way to observe the occurrence of an event is to participate in it, when *A* and *B* interact, third parties who are not part of that joint action are deemed, by assumption, not to know what transpired between *A* and *B*: as far as third parties are concerned, these two actors

---

[51] One reason for the actus reus requirement is that it is impossible for observers to have direct access to an offender's thoughts; the only way for an observer to come to know the offender's state of mind is through an observable projection of it into words or actions. Of course, the observer can be mistaken about how the actions they observe map back into the actor's internal workings.

[52] Cognitive dissonance can lead actors to interpret new information in a manner that is consistent with previous beliefs, and over time, can lead them down a path in which their beliefs diverge significantly from the truth. *See* Manuel A. Utset, *A Model of Time-Inconsistent Misconduct: The Case of Lawyer Misconduct*, 74 FORDHAM L. REV.1319, 1329 (2005) (discussing the role of cognitive dissonance in leading lawyers down a path of moral disintegration and ethical failings).

are a single unit or black box,[53] and their joint behavior, an atomic event.[54] This is equivalent to saying that in order to conclude that a third party, *C*, knows what occurred, it is necessary that, at some later time, *C* interacts with *A*, or *B*, or both; and that this interaction is one that channels to *C* information about the previous one. Recall that under the concurrency approach, a trace is a record of the set of observable interactions by actors in the system; it sets forth the identity of the actors, the type of interaction, and the order in which these occur. Thus, over time, these sequences of interactions will create such a trace, or chain of observable events, accumulating observations in a compositional manner. To see this, suppose that *A* and *B* interacted at time *t*, and (*A* and *B*) and *C*, at time *t + 1*. We can continue to build this chain with the same proviso that each transaction channels only part of the information of previous interactions. For example, at time *t + 2*, [(*A* and *B*) and *C*] and *D* transact, and the latter learns part of what happened in the previous interactions of *A*, *B*, and *C*. One can continue to build this sequence of observations in a similar fashion.

   This compositional approach helps underline that the official evidence set introduced at trial is composed of smaller observation components, all of which bottom out in primitive observation events. Looking at the evidence set through a compositional lens makes it easier to reason about it at different levels of abstraction. For example, a bank employee who steals money and falsifies written and electronic records undertakes various primitive interactions with her environment. The employee will transact with a computer, when she enters the falsified records, and is thus engaged in a joint observation event with that computer, which may later become part of the evidence set. The employee may also engage in an observation event with bank auditors. The employee's interactions with the computer and bank auditors will transfer some information, but also hide some; moreover, when the computer transcript or auditors' testimony is introduced at trial, a new observation event occurs, but one that yields only a partial picture of the earlier interactions.

---

   [53] For example, a computer has a clear interface, including a set of keys, communication ports, screen, and installed programs. When a person interacts with it, the only thing that she can observe is the behavior that is revealed via that interface. The rest of the computer is a black box that prevents her from observing the myriad operations occurring inside. She can interact with internal components only through the parts that have been externalized, or made available to her. This limited set of joint activities between the user and her computer greatly reduce the complexity of using and designing them.

   [54] Suppose that there are three actors, *A*, *B*, and *C*, and that *A* and *B* interact with each other a time *t*; this interaction is an observation event involving *A* and *B*, but not *C*. As far as *C* is concerned this *A-B* interaction is an atomic event; in other words, *A* and *B* together are perceived as a single actor and the only thing that C can know about this black box is what is communicated via an explicit observation event with *C*. Suppose that at time *t + 1*, *C* interacts with *A* and *B*. While all three are now engaged in an observation event, *C*'s knowledge of the *A-B* interaction at time *t* is limited to what *C* can observe through its communication at time *t + 1* with the *A-B* atomic entity.

As this example indicates, the jury is the last observer of an evidence chain that will, by necessity, provide an incomplete picture of the earlier interactions; in other words, while a jury directly interacts with witnesses who were involved in previous interactions, these witnesses can only transfer an incomplete picture of what previously occurred. Assuming otherwise can lead juries to give greater weight to a witness's testimony than that they should. This is because along the way the content of original observations will get transformed—summarized, recorded, verified, combined with other observations, forgotten, or even destroyed. These sequential transformations can create chains of cumulative errors; and collections of relatively minor errors are as able to underwrite a mistaken conviction, as are major ones like witness misidentifications or perjury. At one level, this is an obvious claim; however, my point is that society needs to approach the task of identifying these sources of cumulative errors in a systematic fashion, and that we can learn from the way that engineers have dealt with the same general problem when designing concurrent systems.

3. From Observation Events to Evidence

The first step in analyzing an evidence set is to trace the path that the evidence traversed from the point in time when an offender engaged in a relevant observation event with one or more actors, such as the victim or other witness, to when these observers testify about those interactions. An offender will also transact with its environment through the physical evidence that it leaves behind. Physical evidence involves two observation events. The first is the initial interaction between the offender and recipient, the person or object upon which it is transiently recorded; the second event is the transaction between the recipient and the crime scene investigator who observes and records it. These two steps can be treated as a unit, or as separate observation events—as when we want to determine whether the evidence was planted, degraded, or misconstrued. As the investigation proceeds, other actors will interact with this evidence, which means that keeping an accurate trace of each of these interactions is important; this explains why chain-of-custody rules play an important role in the process.

Using the concept of a trace, we can divide the evidence introduced at trial along a number of dimensions, but I will focus on two of these. The *crime scene trace* is the set of observation events that allow a prosecutor to prove to a jury that a crime occurred. On the other hand, the *defendant trace* is a record of the sequence of observable interactions involving the defendant, beginning at some point before the crime and ending at some point after it. The prosecutor will use the defendant trace to prove to the jury that the person being tried is in fact the one who committed the crime. Wrongful convictions will occur whenever a defendant has left behind a sufficiently rich trace of observable interactions to lead a jury to incorrectly conclude, beyond a reasonable doubt, that the authorities made a correct identification.

The way in which the defendant trace is constructed and verified is critical. Every individual engages in a large number of observable interactions over their

lives, and while these sequences of observed behavior are complex and full of intentional and unintended activity, it is possible for a third party to filter them, extracting slices of a person's history, which they can then use for a variety of purposes. This is exactly what online advertisers routinely do, using sophisticated data mining and classification tools in order to make behavioral predictions about targeted consumers. For example, people leave behind records of observable interactions, capturing not just their purchasing habits, but their education, employment, and relationship with parents, spouses, and children. When a person becomes a suspect, the police and prosecutor will make ample use of the investigatory power granted them, and access large amounts of information recorded over time for a variety of governmental purposes, as well as publicly available information. Moreover, they will sometimes use advanced data mining tools not dissimilar to those used by advertisers.[55]

A defendant's previous observable interactions with the victim—not just threats and arguments, but also the mere fact that they had previously met—may allow a prosecutor to tease out a story, showing motive or opportunity. Even a defendant's previous contacts with the environment in which the crime occurred may become relevant if the defendant "communicated" with that environment by leaving behind fingerprints or DNA evidence. Law enforcement also makes use of a number of scripts to frame investigations, draw inferences, and settle on potential suspects.[56] These include some legitimate ones, such as those based on a person's previous convictions,[57] and statistical data regarding the commission of crimes by family members and acquaintances; they may also include illegitimate scripts, such as those that motivate racial profiling.

4. Some Additional Elaborations

I want to make two further points about the observation events discussed above. First, while an observation event is necessarily a joint activity between the observer and the observed, it may or may not be symmetrical,[58] and can differ as to

---

[55] The ability of computers to retain traces of previous interactions (even if deleted) have made it easier for prosecutors, particularly in white-collar cases, to search through hard drives to extract sequences of emails that can show the progress of criminal activity, cover-up attempts, and so on.

[56] For a discussion on the role of scripts, see Manuel A. Utset, *Reciprocal Fairness, Strategic Behavior & Venture Survival: A Theory of Venture Capital-Financed Firms*, 2002 WIS. L. REV. 45, 88–89 (using concept of scripts from cognitive science to model expected behavior of repeat players in contracting contexts).

[57] While it may be legitimate for the police to use this information, it does not follow that a jury should also have access to it, given well-known biasing problems. Restrictions on the introduction of evidence of previous bad acts can be explained along these same lines.

[58] More specifically, an observation event is a relation O, between actors a and b, such that aOb means that a can observe b's behavior. The relation is symmetrical if both aOb and bOa.

the content or characteristics of the information transferred. When an offender interacts with a victim or a witness, their observations, although part of the same event, may differ; and two witnesses who observe the same crime simultaneously may observe different things, or give conflicting interpretations of the event.[59] In the end, what is important is that whenever one actor observes another they are by necessity involved in a joint activity, even if the second actor is oblivious that it is being observed.[60] For example, a person purchasing drugs from an undercover police officer is able to observe the outward behavior of that officer, but is unable to know what is motivating that behavior or her true identity. On the other hand, the defendant's actions are being observed by the officer as part of an action sequence that will culminate in a crime when the person purchases the drugs; but the officer will be unable to observe that person's state of mind, although a court may conclude that her observable behavior is sufficient to infer intent.

Although one may be tempted to conclude that the observation relation is reflexive—that is, one in which a person is able to observe her own behavior—this conclusion will not always hold. If Peter is absentmindedly whistling, an annoyed Mary may ask him to stop; while Mary is engaged in an observation event with Peter, he may credibly claim that he was not involved in an observation event with himself, telling Mary that he did not realize that he was whistling.[61] One can interpret the voluntariness requirement in criminal law as one that states that a person will be held responsible only if the action sequence that culminated in the harm included a sufficiently meaningful self-observation event to make the person criminally responsible. Time-framing problems in criminal law often revolve around a two part inquiry: first, a determination of the action sequence that the court will conclude is the appropriate one, and second, an inquiry into whether that action sequence included a reflexive observation event that is sufficient to trigger criminal responsibility.[62]

C. "Convincingly" Convicting the Innocent

This subpart examines the conditions under which *two* (or more) potential defendants can be considered to be observationally equivalent and thus legally

---

[59] *See* P Y A Ryan & S A Schneider, *Process Algebra and Non-Interference*, *available at* http://www.citeseer.ist.psu.edu/196401 html (follow "pdf" hyperlink) (describing concurrent systems involving confidential information in which two actors may engage in the same joint event, but both observe different things).

[60] In fact, this approach has been used to model interactions in which messages are passed but the sender can remain anonymous, something that is sometimes important when securing networks. *See generally* Steve Schneider & Abraham Sidiropoulos, CSP and Anonymity (1996) (unpublished manuscript, on file with author).

[61] Similar issues arise when one is trying to explain what it means for a person to take an action. *See* HARRY G. FRANKFURT, THE IMPORTANCE OF WHAT WE CARE ABOUT: PHILOSOPHICAL ESSAYS 47–56, 69–79 (1988).

[62] See Mark Kelman, *Interpretive Construction in the Substantive Criminal Law*, 33 STAN. L. REV. 591, 603–04 (1981).

guilty, even though one or both is factually innocent. In other words, imagine that two defendants are being tried by different juries for the same crime, where each jury believes that they are passing judgment on the sole defendant. The general problem is the following: under what conditions can both juries correctly conclude that both defendants are guilty beyond a reasonable doubt of the same crime? Understanding the role played by observational equivalence in successful, but mistaken, convictions of innocent individuals is important, if for no other reason because it shows how *easily* it can happen—at least in theory; but theory may in fact be the best one can hope for, given the problems associated with empirically testing the system's reliability vis-à-vis wrongful convictions. Nonetheless, to convince policymakers that the problem is potentially systemic, one cannot merely make them aware of the mistaken convictions that have so far come to light (since proportionally their number is relatively small); instead, as Wittgenstein's quote in the epigraph indicates, it is necessary to work in reverse by identifying the road that can lead one from error to the truth.[63]

1. Equivalent Defendants and Mistaken Convictions

How does a prosecutor convince jurors to convict a defendant? What types of proof strategies are available to an ideal prosecutor whose principle goal is to minimize the potential for erroneous results? As I argued in the previous section, criminal prosecutions are built on top of evidence sets that are composed of one or more observation events that are considered legally relevant, and are otherwise admissible. Therefore, whether or not a factually guilty offender is eventually identified, prosecuted, and convicted, and, equally important, whether an innocent offender is mistakenly convicted will depend critically on the observable evidence trace that they each leave behind. Using this evidence, the ideal prosecutor has to prove, beyond a reasonable doubt, that: (1) *someone* undertook an action sequence that violated the law; and (2) that that someone was the defendant.

The first part of this proof strategy is fairly straightforward. Since the identity of the real offender is not known—that is, it still has to be proved—the prosecutor has to introduce evidence that a *hypothetical offender* undertook an action sequence that falls within the equivalence class for the crime in question.[64] If a prosecutor cannot show that someone violated the law, then no one, including the defendant, can be found guilty.

---

[63] My point is not that the evidence is not convincing or that a handful of mistaken convictions would be tolerable; rather it is that to convince skeptics it is often best to show the logical steps by which even relatively small perturbations in the evidence set can lead to convincing, but wrong decisions.

[64] Note that real prosecutors will not necessarily tell jurors that this is what they are doing, or even characterize their behavior in this fashion. Moreover, a prosecutor may combine some or all of these steps when presenting her case. Nonetheless, by analyzing the underlying structure of an *ideal* prosecutor's proof strategy we can identify the types of insidious errors that can creep into the process, which by necessity is the first step in trying to fashion reliable procedural rules.

The second step of the prosecutor's burden is the primary source of mistaken convictions that survive appellate review.[65] The prosecutor has to establish that the defendant and the actual offender are observationally equivalent. However, since the identity of the real offender is not known, the prosecutor must follow a three-step process. First, it has to establish that the behavior of a hypothetical offender is observationally equivalent to that of the *real* offender.[66] Second, it must show that the defendant at trial is in turn observationally equivalent to the hypothetical offender. And finally, by transitivity, it would demonstrate that the defendant is observationally equivalent to the real offender, and thus should be found guilty.

Errors are most likely to creep in during the first step of this tripartite process. For example, suppose that the prosecutor can use the defendant's trace both to: (1) define the hypothetical offender *and* (2) prove that the defendant is equivalent to the hypothetical offender. Then the prosecutor has introduced the potential of bootstrapping itself into a mistaken conviction. In other words, the prosecutor should first have to establish the content of the *real offender's* action sequence, using solely nondefendant-specific observations; this will allow it to specify the hypothetical offender independently of the defendant-specific trace evidence that is not also part of the crime-specific observations. By partitioning the evidence set into the crime trace and the defendant trace subsets, we are able to identify an important source of cross-contamination of the evidence set that can lead to mistaken convictions.

2. *The Advantages of the Observation Event Approach*

Characterizing the evidence introduced in trials as based on a sequence of observation events helps us accomplish five things. First, it allows us to analyze the official evidence sets in trials in a uniform manner, making it easier to identify contexts in which gaps in the evidence and erroneous observations can lead to mistaken convictions. Second, it clearly distinguishes between evidence of internal nonobservable events and external ones, which ultimately are the only ones a witness can observe. Third, it helps underscore the important fact that both the identity of observers and the context in which observation events occur matter. For example, a court's willingness to accept evidence from a joint observation event

---

[65] An innocent person may be mistakenly convicted because a judge has misapplied the law by improperly extending the contents of a crime's equivalence class—that is, by concluding erroneously that the defendant's action sequence triggers crime *C*. But a mistake regarding the equivalence class is more likely to be corrected by an appellate court. As a result, my principal focus is on contexts in which an innocent person may be mistakenly convicted because they were considered to be equivalent or identical to the actual offender, or where the defendant is in fact the offender, but is able to raise reasonable doubt that they are equivalent to the hypothetical offender.

[66] The hypothetical offender can be seen as a sort of specification and a defendant as a possible implementation of it; at the very least a prosecutor should have to prove to the jury that convicting the defendant is a reliable step—that is, one that is almost guaranteed to be correct (beyond a reasonable doubt).

will often depend on the identities of the parties involved, as is the case with evidence rules on hearsay, privileged communications, and business records.[67] Fourth, the concept of a joint observation event can be extended to the relationship between a witness, on the one hand, and the police or the jury, on the other hand, making it easier to identify each subsequent interaction that can cause a witness's observations to go astray. Fifth, the approach's compositional nature makes it easier to analyze various ways in which different observation events can be combined and how the process can affect the meaning of the composite observation event.

D. Over-Inclusive Equivalence Relations

I argued that a prosecutor can convict a defendant by proving to a jury that the defendant is observationally equivalent to the real offender, something that ideally (from a mistake-reducing perspective) will involve an intermediate step in which the prosecutor proves that the defendant and real offender are in turn equivalent to a hypothetical offender. The best possible result is one in which the real offender is the only person who is in an equivalence relation with the hypothetical offender. However, this is impossible to achieve in practice, and thus policymakers need to consider the extent to which procedural rules affect the content and size of the offender-equivalence class. This is because, under my model, every member of that class is equivalently guilty of the same crime. As a result, all other things being equal, the greater the number of individuals who possess observable action sequences that are equivalent to the hypothetical offender's, the greater the potential for mistaken convictions. In other words, the weaker the exclusion property of the relation, the more likely that among the myriad action sequences that people possess, one will be deemed to be a sufficiently good fit to the hypothesized crime sequence.

One general approach for testing the stability of an equivalence relation is to apply the same operation or transformation to two members of an equivalence class and see if they are still equivalent afterwards; if they are, we can say that they are congruent, a deeper relation than mere equivalence.[68] To see this, suppose that a and b are equivalent; if a is transformed into a' using transformation T and b is

---

[67] *See* FED. R. EVID. 501, 803(6).

[68] More precisely, a group (G, *) is an algebraic structure that associates a set S with a binary operation * that satisfies four general axioms, including closure (for all a, b members of G, the result of ($a * b$) is also in G), associativity (for all a, b, c in G, ($a * b$) * $c = a * (b * c)$)), as well as the existence of an identity element and an inverse element. A congruence is a stronger form of equivalence—one preserved by operation on the set. If a, a', b, and b' are members of group (G, *) and ~ is a binary relation on set G, then we can say that a and b are not only equivalent but also congruent if after applying the operation * to a and b the resulting members of G, a', and b' are also equivalent to each other. In other words, we can say that ~ is not only an equivalence relation but also a congruence whenever the following holds: if (a ~ b, $a * c = a'$, and $b * c = b'$), then (a' ~ b' or $a' \equiv b'$). *See* DURBIN, *supra* note 35, at 64–65.

transformed into *b´* using the same transformation, then *a´* and *b´* will either continue to be equivalent, or they will not. To the extent that they remain equivalent, and if that "test"[69] was well chosen for the task at hand, then that transformation has produced new information about the relationship between *a* and *b*: that their equivalence is strong enough to survive that test; or, equivalently, that their "sameness" is invariant under that transformation. The greater the number of transformations that leave invariant the equivalence relation between *a* and *b*, the deeper or more robust that relation.[70]

How do we apply these tests or transformations to distinguish or equate two individuals? What types of transformations will allow us to check the robustness of an equivalence relation between persons *M* and *N*? Assume that a person *M*, who is born on year *y* undergoes a number of changes or transformations between years *y* and *y + 40*. One way to model these changes is to posit that each time that *M* undergoes such a change, she becomes a slightly different person: for example, transformation one through ten, turns *M* into M^1, $M^2 \ldots M^{10}$. An important problem in the philosophy of identity is the question of determining the extent to which one can say that person M^{10} is the same as *M*. Some transformations will leave *M* invariant—*M*'s name, parents, eye color; others will not—*M*'s age, height, and knowledge. In the same manner, imagine that *M* and *N* are part of the same equivalence class, and that we have chosen a set of transformations T^1 through T^{10} that are appropriate and effective for the particular question at hand.[71] One would then apply these transformations to *M* and *N* to produce new sets of *M*s and *N*s, and then determine whether these new instantiations continue to preserve the equivalence relation—that is, whether or not $M^1 \sim N^1$, $M^2 \sim N^2$, and so on.[72]

Society can use this "transformation" method to separate guilty offenders from innocent individuals. While some procedural rules already make use of this sort of device, my main point is that policymakers and commentators should give greater attention to the task of systematically designing and using equivalence-breaking or invariance tests. Ultimately, the difficult problem is that of identifying the most effective transformations to produce the level of reliability demanded of procedural rules. Importantly, however, the aim is not necessarily to make the

---

[69] In this Article the terms transformation, operation, and test mean the same thing.

[70] *See* NOZICK, *supra* note 33, at 75–83 (discussing a similar approach to determine the level of objectivity).

[71] This is not an easy task, given that one needs to choose an independent criterion for testing the transformations themselves. *See* NOZICK, *supra* note 33, at 79–83 (discussing the general problems of choosing transformations and suggesting a recursive mechanism for testing and modifying them). To the extent that it is an area in which the testers are repeat players, then one would expect that over time they will develop a sense of what types of transformations are effective.

[72] Note that I am bracketing a number of important issues that are beyond the general scope of this paper, such as a number of problems that can arise when choosing, using, and testing transformations, including the fact that the order in which the transformations are applied may lead to cycles if transitivity is not preserved. This should not affect the general arguments made here.

equivalence class as small as possible; after all, one can imagine certain transformations that will reduce it from a size of $n + 1$ to size n, where that last transformation excludes the real offender, and, thus, increases the probability that an innocent person is convicted instead; for example, if one of the transformations is based on factors such as religious beliefs or race that do not sufficiently track the real crime.

IV. IMPLICATIONS OF THE OBSERVATIONAL EQUIVALENCE APPROACH

I begin this Part by setting forth two counter-intuitive results that follow from the observational equivalence theory of wrongful convictions developed in Part III. In the first subpart, I draw a distinction between fine-grained and coarse identifications, where an identification is more fine-grained the greater the number of defining characteristics of an offender that an eyewitness uses to identify her. I argue that jurors will give greater weight to fine-grained *misidentifications* than to coarse ones, and, thus, that all other things being equal, once a misidentification has occurred, fine-grained ones are more likely to lead to wrongful convictions. I then analyze one type of coarse misidentification—cross-racial misidentifications. I show that one can reduce the likelihood of mistaken convictions based on this sort of eyewitness mistake by making use of a witness's very inability to draw fine-grained distinctions between individuals of another race.

In the second subpart, I turn to the question of exonerations. I show that the fact that it is impossible to identify all wrongfully convicted individuals at the same time increases the likelihood that innocent prisoners, behind a veil of ignorance, would choose a procedural rule that would keep some wrongfully convicted individuals imprisoned, even when one can prove their innocence.

In the rest of Part IV, I develop some further implications of the observational equivalence model. I first look at a number of problems related to the way that crime equivalence classes are specified and modified. Finally, I argue that offenders owe a moral duty to wrongfully convicted individuals and that there is a deep tension between two societal goals in the criminal law area: (1) protecting personal autonomy; and (2) preventing wrongful convictions.

A. Cross-Racial Misidentifications:
When Two Equivalence Mistakes Can Make a Right

Eyewitness misidentifications are an important cause of mistaken convictions.[73] The observational equivalence approach suggests that misidentifications are a problem whenever they make it easier for a witness and jury to conclude that an innocent defendant is equivalent to the perpetrator.

---

[73] *See, e.g.*, Gross, *supra* note 4, at 18.

1. Identification

Witness identifications are an important type of observation event, since they create a direct link between the defendant and the offender. I will assume that an eyewitness identifies a suspect in a point-wise manner: the witness will first isolate a set of defining characteristics that she can later use to pick-out the offender from a pool of suspects; she will do the latter by checking whether any of the suspects has a sufficient number of equivalent defining characteristics to allow her to make a positive identification.[74] Under this approach, a witness is correct with regard to a defining characteristic whenever the suspect and offender have the same characteristic, and incorrect in all other cases.

A misidentification can occur in two general types of scenarios. The first are fine-grained misidentifications: those in which a witness makes an identification based on a large number of observation points, but makes a mistake about a sufficient number of them. The second type involves coarse misidentifications, which are based on very few defining characteristics. Some witnesses resort to coarse identifications because, at the time of their initial observation, they were unable to fully perceive an offender's defining features; for example, when events occurred quickly or in bad lighting. Others do so even though they were able to get a good look at the offender. This can occur when a witness is unable to make fine distinctions between defining characteristics, as is sometimes the case when people make cross-racial identifications. As we will see below, this means that one way to address the problem is to adopt procedural safeguards that weed out coarse identifications.

2. Fine-Grained Misidentifications

Let us begin with fine-grained misidentifications, those that occur even though a witness makes use of a large number observation points. As a general matter, one would expect that the greater the number of data points, the less likely that the witness will reach the wrong conclusion. For example, a witness may not remember an offender's hair color, but recalls his height, body type, and facial features. The witness can use these to compensate for her observational gap regarding hair color. This is analogous to an identification made by multiple eyewitnesses, which, as a general matter, is a stronger one than one made by a single witness.

However, multiple eyewitnesses can, in certain instances, make matters worse. If they all share the same observational shortcoming, if they are mistaken along the same dimension, multiple eyewitnesses can make misidentifications that

---

[74] The order in which a witness takes these steps can affect the outcome. A witness is more likely to make a mistake if she selects the relevant characteristics after she observes the suspect and only then tries to see if she can map them onto the person she saw commit the crime. One way to assure the correct ordering is to require that a witness describes the offender—for example, to a sketch artist—before she is exposed to the pool of suspects.

produce greater harm than one by a single witness. This is because a jury will give greater weight to their reinforced mistaken identification. By analogy, one would expect that a jury will give greater credence to the testimony of an eyewitness who can testify about multiple defining characteristics than one who, as in the cross-racial context, is unable to provide a fine-grained identification. It is this added weight that makes it more difficult for a truly innocent defendant to challenge the misidentification.

Finally, one can generalize this argument to misidentifications based on DNA evidence. Juries give great weight to DNA evidence for the very reason that if two DNA samples are deemed equivalent, the probability that they came from different individuals is infinitesimally small. This means that if the defendant's DNA is found at the crime scene, it is extremely unlikely that the DNA came from someone else. However, it does not follow that the defendant necessarily left the DNA behind while he was committing the crime in question. As a result, one would expect that DNA evidence will increase potential mistaken convictions whenever a jury gives undue weight to the invariance condition connecting the DNA to that person, and not enough weight to potential variance conditions—for example, the fact that an innocent defendant's DNA could have been left behind at the crime scene for some other reason.

3. Cross-Racial Misidentifications

A cross-racial identification based on a small number of defining characteristics is unlikely to carry the same rhetorical force as a fine-grained identification. As a result, all other things being equal, it should be easier to challenge the reliability of such a coarse, cross-racial identification. Moreover, as I will now argue, society can exploit a witness's inability to make cross-racial distinctions to reduce the probability of wrongful convictions due to cross-racial misidentifications; in short, two misidentifications are better than one. The motivating intuition is the following: a witness who has difficulty making cross-racial identifications will have the same problem in *all* contexts in which she must identify a person of a different race. Such a witness should have an equally hard time making fine-grained distinctions in subsequent observations—for example, a lineup, where she will have to pick-out the offender from a pool of racially homogeneous suspects.

I argued in Part III that criminal prosecutions involve showing that (1) a crime occurred and (2) identifying the offender;[75] cross-racial misidentifications can only occur along this second dimension.[76] I also argued that observation events are time

---

[75] *See supra* Part III.C.1. (arguing that an ideal prosecutor will use a two-step proof strategy in which she first shows that someone—a hypothetical offender—committed a crime; and then uses this fact to convince the jury that the defendant at trial is equivalent to this hypothetical offender; and by transitivity, is equivalent to the actual offender).

[76] While a person's status may play a role in determining whether a crime occurred, as when a fiduciary or trustee is involved, this is a different question than that of identifying the actual offender. It is easy to determine a defendant's status ex post; but the real

sensitive: a person who witnesses someone committing a crime is engaged in a joint event with the perpetrator; but this is just the first of a sequence of observation events regarding that interaction between the offender and witness.[77] Before the witness can transfer any information to a jury, she will have to interact with the police and prosecutors, and with potential suspects—for example, at the time of a lineup and at trial.

I will assume the worst possible case, one in which an eyewitness of race X is wholly unable to distinguish between any two individuals of race Y—that is, she cannot tell them apart using defining features and characteristics. Suppose that this witness observes a person of race Y commit a crime and is asked to provide a description to a sketch artist. The artist will ask the witness to choose from a set of sample features—eyes, hair, mouths, and so on. Since, by assumption, the witness is unable to distinguish between individuals of race Y, she would have to randomly choose a set of features. This will of course reduce the probability that the police will find the right person, or, in some instances, anybody who fits the description. Assume, however, that the police have arrested a potential suspect. The witness will now have to pick that suspect from a lineup. Again, by assumption, the witness is unable to distinguish between the individuals in the lineup, all of which are of race Y. This means that the witness will either choose one at random or conclude that the person she saw commit the crime is not in the lineup.

Suppose that a benevolent lawmaker is concerned about cross-racial misidentifications and wants to adopt identification mechanisms that address the problem; at the same time, she wants to make sure that these do not adversely affect witnesses making fine-grained identifications. I argue that one possible solution is to increase the number of choices available to witnesses, both at the time of giving a description and when choosing from a lineup.

To see how such a "Where's Waldo"-type solution would work, suppose that sketch artists are required to show witnesses a large number of types of eyes, mouths, noses, ears, and so on before a witness can settle on any one of them. Increasing the number of defining features will not affect witnesses who are able to make perfectly fine-grained identifications. On the other hand, the witness who can only make wholly coarse identifications, will, as we saw above, choose features at random. Now suppose that after a witness has chosen a set of features, a number of sketches are generated, one with the chosen features, and the rest with randomly generated ones. If the witness fails to identify the correct sketch, the police can conclude that the one she chose is not a reliable one.

The police can use analogous procedural safeguards at the time of a lineup. Suppose that a sketch is deemed reliable and that a suspect is arrested.

---

question is whether the defendant undertook an action sequence that is equivalent to one in that crime's equivalence class. *See supra* Part III.A. (arguing that crimes are completely defined by a set of equivalent action sequences any one of which is sufficient to trigger a violation of the law).

[77] *See supra* Part III.B. (regarding the construction of evidence sets through a sequential composition of observation events).

Additionally, suppose that a witness is not allowed to review the sketch in subsequent periods. Instead of the standard lineup with five or six possible suspects, the lawmaker can require the police to use lineups with a much larger number of suspects. This can be accomplished through a single, large lineup, or a series of smaller ones, where the witness is not allowed to make a final judgment until she has seen the complete array of suspects. Since only one of the suspects is the one that the police arrested, increasing the number of possible choices will decrease the probability of a lucky identification by a witness using a coarse screening mechanism. Exposing a witness to a large array or suspects will not, by definition, affect the witness making a perfectly fine-grained identification.

Most witnesses will fall somewhere in between the two extremes of the perfect and challenged witnesses. However, this does not affect the general approach. The basic intuition is that eyewitness identifications should remain invariant when one increases the sets of distinguishing characteristics and possible suspects; moreover, the identifications that a witness makes in different time periods and contexts, should also be invariant. If witness identifications are unable to survive these two invariance conditions, the state should suspect their reliability.

B. Should Innocence Lead to Exoneration?

Commentators and policymakers have assumed that all innocent individuals who are wrongfully convicted are equivalent along two dimensions: (1) they are all in prison for an illegitimate reason; and (2) if it is determined with sufficient certainty that they are innocent, their convictions should be reversed, and to the extent that this was the sole reason that they were in prison, they should be freed. I now want to argue that this second statement does not necessarily follow from the first. In other words, there are cases in which a person is in prison for an illegitimate reason but releasing them immediately may require additional justification. My point here is not that we should leave innocent individuals in prison; rather it is to show why using categorical equivalence relations can hide numerous subtleties, and ultimately undermine the more general goal of minimizing the number of wrongfully convicted individuals in prison.

Suppose that society is considering adopting the following very plausible exoneration rule: whenever it is determined with one hundred percent certainty that a prisoner is innocent, she should be immediately exonerated and released. My claim is that if they were asked to adopt or reject this rule, a group of hypothetical innocent offenders, residing behind a veil of ignorance, would reject it; in short, they would choose to keep some innocent individuals imprisoned. Why? First, note that if it were possible to identify all current innocent prisoners concurrently and release them at the same time, the hypothetical prisoners would favor the immediate-exoneration rule. However, in the real world, the state cannot identify innocent prisoners concurrently but must do so sequentially, which means that the order in which they are identified and released will matter. In particular, if the innocent prisoners behind the veil of ignorance do not know at what point in that sequence society will learn about their wrongful convictions, they will reject the

rule unless they were convinced that those who are released before them will not reduce the likelihood of their own future exoneration.

For example, a recidivist whose most recent conviction is overturned and offends again will impose a negative externality on the innocent still in prison. This means that judges and lawmakers who observe the recidivist's new crime will have a greater incentive to increase the burden faced by subsequent prisoners seeking exoneration. These heightened hurdles matter because barring clear DNA evidence, proving innocence with one hundred percent certainty is usually impossible. As a result, even when an elected prosecutor or judge is convinced that a prisoner is likely innocent, she may hesitate to exonerate him if she is not convinced that he will not reoffend.

Finally, it is not only other innocent prisoners, judges, and prosecutors who are affected by the problem identified in this subpart. An organization dedicated to achieving the exoneration of the innocent that is required to raise funds from public and private sources will, at least in theory, have an incentive to take on potential clients in an order that takes into account the probability that an exonerated prisoner will re-offend. In conclusion, maximizing the number of innocent prisoners who are identified and released may require taking into account their "average guilt," by taking into account the crimes that they have already committed, since this will provide some indication of the probability that they will reoffend; the probability that they will impose a negative externality on the innocent still waiting to be discovered.

C. *Defining Crime-Equivalence Classes*

As we saw in Part III, each action sequence that triggers a violation of crime *C* can be considered a member of its equivalence class. How would one go about determining these members? One way is to use the following bottom-up and top-down approaches. Suppose that ten individuals have been convicted of crime *C*. By examining the evidence in those cases one should be able to extract the relevant set of actions that led to their convictions.[78] These ten action sequences are equivalent in the sense that they all triggered violations of crime *C*. One could engage in the same exercise with those cases in which defendants were acquitted.[79]

---

[78] In other words one would examine the set of actions that were observable by the jury such that they were able to conclude that the defendant was in fact the person who committed the crime. I do not mean to suggest that a jury's decisions will not involve judgment calls as to the truthfulness of different testimony or different interpretations of the evidence. However, one should be able to extract a set of observable events (while abstracting away from other factors) that is the smallest set that could have triggered the conviction. Among other things, this would allow one to compare different laws to determine how well-specified (or not) they are.

[79] One reason for using both a bottom-up and top-down approach is to correct for the potential hindsight bias in the bottom-up investigation—that is, some actions may be deemed the most relevant just because they were the ones that occurred and thus are most

In addition to this bottom-up approach, one can also imagine a top-down approach. The first step would be to identify at least one hypothetical action sequence that would trigger crime *C*. This does not necessarily need to be the most basic, essential, or paradigmatic one.[80] This first action sequence is, by definition, equivalent to itself. The next step is to transform it by adding or removing actions, reordering them, or even including an error interval that captures potential errors that a jury may make when it tries to fit the defendant's behavior within the definition of the crime.

Under these two approaches the goal is to identify the observable action sequences that are part of crime *C*'s equivalence class, by looking at: (1) the evidence of the observable action sequences that have been found to fall within that class, or (2) identifying a set of contingent members of the equivalence class and an accompanying set of admissible transformations. As with other equivalence scenarios it may be possible to reach different results by using a different set of transformations. Nonetheless, under this approach, a person who wants to define or modify the equivalence class for a crime will have to both identify and justify the transformations that they are using. Transparency is always the first step in dissolving indeterminacy.[81]

D. Unstable Crime Definitions

In an ideal world, the state would clearly specify the set of action sequences that will trigger criminal responsibility and make this information public so that those who are within a crime's scope can plan accordingly. This is usually impractical given that it is impossible for lawmakers to fully anticipate all of the contexts in which liability for a specific crime should attach. As a result, the size and content of a crime's equivalence class will change over time as courts transform it in legally valid ways—i.e., through interpretations that add or subtract action sequences.[82] Publicity problems can arise if these transformations occur after an alleged crime.

More generally, the act of committing a crime is a concurrent or joint activity between an offender and the state, given that a person cannot violate a law unless

---

salient. For a discussion of the hindsight bias, see Jeffrey J. Rachlinski, *A Positive Psychological Theory of Judging in Hindsight*, 65 U. CHI. L. REV. 571 (1998).

[80] In fact, picking an action sequence as the paradigmatic one can produce an anchoring effect that will distort the way that it is transformed into other sequences. This may of course be the desired effect in some cases, but it is not necessarily so. For an analysis of the anchoring effect, see Greg Pogarsky & Linda Babcock, *Damage Caps, Motivated Anchoring, and Bargaining Impasse*, 30 J. LEGAL STUD. 143 (2001).

[81] In fact as we saw above, concurrency theory associates the concept of indeterminacy with that of nonobservability. *See* Part.II.D.

[82] *See, e.g.*, H.L.A. HART, THE CONCEPT OF LAW 97–107 (1961) (discussing rule of recognition).

that law is in existence at the time;[83] moreover, the actor must be a member of the class of individuals within the law's scope.[84] Although the problem of legal indeterminacy is generally attributed to the inability of lawmakers to anticipate all contexts in which a law should apply, the actual problem is a slightly different one. An unstable legal rule L_1 is one that can be changed into a "new" rule L_2 by applying a legally admissible transformation to the set of action sequences that fall under its equivalence class. A rule's indeterminacy will thus be a function of the manner in which the set of legal transformations are defined and publicized. A decision maker subject to law L_1 does not need to know the exact parameters of law L_2, only the set of operations that a court can apply to transform L_1 into L_2.

For example, there are some acts, such as killing someone or forcing them to engage in nonconsensual sex, where the illegality is communicated via other methods, such as moral strictures.[85] This is true even though it is highly unlikely that offenders will know the exact language or parameters of the prohibited behavior. The effectiveness of these indirect communications depends on a variety of factors, not the least of which is the offender's capacity to make the requisite inferences. However, there may be some transformations that will not be as cognizable. For example, a person may be aware that shooting and killing another person will bring her under the scope of one or more legal rules, but may be unaware of when she is "legally" permitted to shoot and kill an intruder or abusive spouse. Similarly, a person may not be able to easily infer the full panoply of statements and behavior that can transform consensual sexual activity into an instance of rape.

---

[83] Ex post facto laws, which are prohibited under the Constitution, are exactly those in which the state fails to meet this concurrency condition by charging someone for behavior that became illegal only after the fact. *See* U.S. CONST. art. I, §§ 9, 10; Cal. Dep't of Corr. v. Morales, 514 U.S. 499, 504 (1995); Collins v. Youngblood, 997 U.S. 37, 41–42 (1990).

[84] Importantly, the set of offenders that fall within a law's scope either has to be empty or contain more than one actor. A law that singles out one offender is a bill of attainder, which is prohibited. *See* U.S. CONST. art. I, § 9; *see also* United States v. Brown, 381 U.S. 437, 440 (1985) ("the Bill of Attainder Clause was intended not as a narrow, technical . . . prohibition, but rather as an implementation of the general separation of powers, a general safeguard against legislative exercise of the judicial function.").

[85] Such acts are malum in se as opposed to malum prohibitum. *See, e.g.*, Kempe v. United States, 151 F.2d 680, 688 (8th Cir. 1945) ("Crimes have been divided according to their nature into crimes mala in se, and crimes mala prohibita; the former class comprises those acts which are immoral or wrong in themselves, such as murder, rape, arson, burglary, and larceny, breach of the peace, forgery, and the like, while the latter embraces those things which are prohibited by statute because they infringe upon the rights of others, although no moral turpitude may attach . . . ." (quoting 22 C.J.S. *Criminal Law* § 8 (1944))).

E. Equality and the Single Jeopardy Condition

A general principle of criminal justice is that "equals should be treated equally": a law should be applied equally to two independent, but otherwise identical criminal acts, committed by two identical individuals. But the actual meaning of equally and identical vary. If twins commit the same crime under identical circumstances, they both should be subject to the same treatment. However, barring twins and identical instances of the crime, how different do two individuals have to be before we can conclude that they are not equivalent? Does equivalence under the law mean that everybody should be equally likely to be wrongfully convicted? If we are equivalent as to guilt should we not also be equivalent as to innocence? In other words, if individuals of a certain race or socio-economic class are more likely to be wrongfully convicted, we may want to ask for a justification for such a state of affairs.

There is a second related principle of criminal liability, although it is one so axiomatic that it rarely receives attention or is clearly specified: if person M is factually guilty of single-handedly committing crime C, then N is necessarily factually innocent vis-à-vis crime C. It follows that M and N cannot both be convicted for committing crime C. More specifically, they both cannot be convicted concurrently; they may be convicted sequentially; that is, if M is convicted first, then N cannot be convicted of the same crime unless M's conviction has been overturned.[86] If $twin_1$ commits crime C, and $twin_2$ had nothing to do with that crime, then only one of the twins can be convicted. If the only evidence is a witness's identification, but the witness is unable to distinguish between the twins, then the law cannot convict both.[87] Barring some independent observable factor to distinguish the two, it should in fact be impossible to convict either, since evidence about the existence of the indistinguishable twin should be enough to raise reasonable doubt of guilt. We can call this the "single-jeopardy" condition.

Society's ability to effectively implement equality under the law and the single jeopardy conditions will depend in large part on the ability of designers and users of substantive and procedural legal rules to make sense of the concept of equivalence. The single jeopardy condition only requires that if a crime is

---

[86] They both may be convicted concurrently if there is not only a mistake as to identity (whether M or N was the factually guilty party), but a mistake as to the number of individuals involved in the crime (whether either M or N was single-handedly involved, or both M and N were involved).

[87] If for some unrelated reason $twin_2$ feels morally responsible for $twin_1$'s actions and agrees to stand trial and is convicted, should we consider her equivalent to other factually innocent individuals in prison? Is this different from a defendant who is (1) innocent of crime C_1 (for which she was arrested) but factually guilty of the yet undiscovered crime C_2, and (2) enters a plea bargain for C_1 to avoid further investigations that may lead the authorities to discover the second crime? These are difficult questions with no obvious answer, unless one has a theory of whether all factually innocent convicts should be treated equally, and, if not how one would distinguish between them.

committed by one person, then only one person can be convicted; it says nothing about *who* can be convicted. This condition has an important effect on individuals who have been mistakenly convicted. This is because it requires that only one person can be guilty of a crime at any one time; if person *M* has already been convicted for a crime that was in fact committed by person *N*, then the police will cease searching for *N*. Additionally, the best way for an innocent person to have their conviction overturned is if the actual offender is caught and prosecuted, and the latter cannot occur until the innocent person's conviction has been reversed.

F. An Offender's Responsibility to Innocent Suspects

When offenders commit crimes, they want to avoid detection and punishment. To lower the probability of detection, they may undertake avoidance activities aimed at assuring that the evidence trace they leave behind, memorializing their observable behavior, is insufficient to show that: (1) a crime occurred; or (2) if the crime is detected, that they are the offender.[88] One type of avoidance strategy available to an offender is to take one or more actions that will increase the probability that an innocent person will be mistakenly identified as the perpetrator. If an offender's avoidance activity leads to an innocent person's wrongful conviction, is that offender morally or legally responsible? Should they be? Committing a crime separates the offender from everyone else in society: it identifies her with that crime and opens her to punishment and moral opprobrium. An offender is responsible for the harm borne by her victim; but one can make a strong argument that she is also responsible for the harm borne by the innocent person who is mistakenly convicted. This is true even if that person is acquitted or never brought before a jury, given the various costs borne by those suspected of crimes.[89]

---

[88] The surest way to accomplish this is to leave absolutely no observable evidence trace. But, this is too strong a condition; an offender can escape liability as long as her observable action sequence is sufficiently close to a legal one, such that a third-party observer cannot tell the two apart. As we have seen, the state will be unable to show that a crime occurred, unless it can prove that an offender's observable action sequence is equivalent to one in a crime's equivalence class. Computer scientists working on the question of system security adopt a definition of confidentiality based on the absence of information flow, which is based on a similar type of argument. *See* SCHNEIDER, *supra* note 13, at 7–8 (arguing that in order to show that information does not flow from *A* to *B* one only needs to show that *A*'s actions do not affect *B*'s observations).

[89] For example, the state pierces the realm of privacy or autonomy of people who are wrongfully identified as the perpetrator of a crime. This produces losses that are internally tangible, even if not observable or measurable by others; it may lead to more tractable losses, such as attorney's fees and reputational losses, which like those borne by victims of slander will not necessarily be reversed when names are cleared. *See, e.g.*, Jewell v. NYP Holdings, Inc., 23 F. Supp. 2d 348, 361–62 (S.D.N.Y. 1998); Eric Harrison, *Security Guard Cleared in Olympic Comb Case*, L.A. TIMES, Oct. 27, 1996, at A1, A20. In fact, a nomenclature has developed to try to deal with this reputational problem—for example,

It is possible to envision a stepwise process by which the offender's moral responsibility increases with each procedural hurdle that the innocent defendant clears on her way to being wrongfully convicted. It short, while an offender's decision to commit a crime imposes a direct harm to the victim, her decision to avoid detection imposes a second-order expected harm on potential innocent suspects.[90] Moreover, there are a number of very plausible scenarios in which this second-order harm is greater than the actual harm to the victim. For example, an innocent person who is mistakenly convicted and imprisoned for a burglary of an empty home will no doubt be harmed by a greater amount than the actual victim.

When an innocent person is wrongfully convicted, the state is undoubtedly responsible; but the state is the last cause in a causation chain that begins when an offender violates the law. The offender may ultimately be absolved of at least some moral responsibility if the state's role as an intervening cause is sufficiently wrongful or reckless.[91] But regardless of how harm and desert are distributed between the parties, the general problem of guilt and innocence bottoms-out in a tripartite relation between the state, the actual offender, and the innocent convict.

This tripartite relationship raises important questions relating to the intersection of identity and autonomy. Personal identity is a necessary condition for autonomy: two people cannot be autonomous vis-à-vis each other, unless each knows where the other begins and ends. But autonomy—our ability to keep things (such as our behavior) internal to ourselves, away from the gaze of others—is a central component of identification mistakes in criminal law. When a perpetrator and an innocent person are observationally identical defendants, their personal identity will be extinguished for questions of criminal responsibility and punishment. In short, observational autonomy helps foster our identity and ability to act freely; but, at the same time, it can undermine that very identity and freedom.[92]

---

calling a suspect a person of interest or even a material witness, although the latter may be used to meet certain procedural constraints.

[90] This expected harm will depend on the probability that the authorities will investigate and try one or more innocent suspects and on the magnitude of that harm.

[91] Causation chains of this sort play an important role in both torts and criminal law and raise a number of difficult issues regarding how liability should be distributed among the various parties whose actions bear on causation. *See, e.g.*, H. L. A. HART & TONY HONORÉ, CAUSATION IN THE LAW 235–53 (1959) (outlining liability and causation via foreseeability).

[92] A caveat—the reason to tease out this internal tension between observational and behavioral autonomy is to better understand the nature of the tripartite relationship between the state as observer and the factually guilty and innocent individuals who are to be distinguished; it is not (at this level of generality) to make any claim about the parameters of autonomy and privacy within the criminal justice system.

V. CONCLUSION

My goal in this Article has been to develop a model of wrongful convictions that can account for cumulative mistakes that, when taken together, can convince a jury that an innocent person is in fact guilty. I argued that to reduce the incidence of wrongful convictions it is necessary to understand how these cumulative errors can arise and know how to identify them in real time. To do this, the model reduces the criminal justice process to sequences of observable behavior by offenders, witnesses, the police, prosecutors, and innocent individuals. It also defines a crime by the set of equivalent action sequences, any of which is sufficient to trigger a legal violation; these sequences constitute a crime's equivalence class.

At trial, a prosecutor will introduce an evidence set, composed of observation events, and containing one or more action sequences; she will use it to prove to a jury that: (1) someone undertook an action sequence that falls within the equivalence class of the crime in question; and (2) the defendant was that person. Under the model, wrongful convictions will occur whenever an innocent person has undertaken a sequence of actions that is sufficient to allow a prosecutor to convince a jury of these two things, beyond a reasonable doubt. I showed that it is possible for more than one person to be equivalently guilty of the same crime; while they cannot both be convicted at the same time, it is possible to convict both sequentially.

Additionally, I described two counterintuitive results surrounding wrongful convictions. The first one involves wrongful convictions due to cross-racial misidentifications. I showed that it is possible to reduce these by adopting mechanisms that exploit a witness's very inability to make fine-grained identifications of individuals of other races. The second result follows from the fact that it is impossible to identify and release all innocent prisoners at the same time. I argued that the fact that these innocent prisoners will be identified sequentially, if at all, means that the order in which they are released matters: an exonerated prisoner who reoffends will impose a negative externality on the innocent individuals still incarcerated. As a result, I argued that hypothetical innocent prisoners, behind a veil of ignorance, would reject the following exoneration rule: any prisoner who is shown to be innocent with one hundred percent certainty should be immediately released.