The Human Factor: Algorithms, Dissenters, and Detention in Immigration Enforcement

Robert Koulish[b] **and Ernesto Calvo**[a,b,1]

[a]iLCSS; [b]GVPT-University of Maryland

April 1, 2019

This article examines changes in the punitive bias of Immigration and Customs Enforcement (ICE) officers in the 2012 through 2016 period. We define punitive biases as the officers' higher likelihood of expressing a written dissent with Low Risk Classifications cases when compared to high risk classification ones. We provide evidence that contextual factors enter into the immigration decision process in two different ways: one, by shifting the officers' frame of references based on the number of cases, the compounded risk portfolio available in processing centers (e.g. the caseloads flow), and in response to external shocks, such as elections. Second, by allowing policy preferences to edit the risk algorithm (i.e. inserting, deleting, and reweighting the importance given to different types of violations). The analysis draws data from 1.4 million immigration detention risk classification assessment cases between 2012 and 2016 received pursuant to the Freedom of Information Act (FOIA).

Immigration | electoral cycle | punitive bias | risk classification | immigration detention.

When Crane v Napolitano was filed on behalf of the Immigration and Customs Enforcement (ICE) agents against the Obama administration in 2012, little did people realize the influence it would have over immigration detention in the next four years. The same with the Council's decision to endorse Donald Trump during the 2016 presidential campaign. The legal case and the political endorsement provide evidence of ICE's antipathy towards the Obama Administration during elections. They also highlight the enforcement priorities and the political preferences of ICE agents themselves.

This article examines the tension between risk algorithms and end users (officers and supervisors) in immigration detention decisions (detain/release).[1] Much scholarship on actuarial risk endorses the view that risk is about objectivity, efficiency, and accuracy in decision-making structures.[2] This belief is a cornerstone of Immigration and Customs Enforcement (ICE) decisions to detain or release immigrants and categorize custody classifications. Thereby, since 2012, ICE has deployed risk classification assessment (RCA) to lend objectivity, efficiency and accuracy to immigration detention. Our research shows that contextual factors affect the implementation of this algorithm as well as the algorithm editing decisions in immigration detention.

We provide evidence that contextual factors enter into the immigration decision process in two different ways: one, by shifting the officers' frame of references on risks assessments with increases in the number of cases, changes in the risk portfolio in ICE's processing centers (e.g. the caseloads flow), and in response to external shocks such as elections. Second, by allowing policy preferences to edit the risk algorithm, by inserting, deleting, and reweighting the importance given to different types of violations. The process of algorithm editing, we argue, induced changes in the formal rules that allocate rewards and punishments. Over time, accommodation between the preferences of end users (officers and supervisors) and the algorithm toggled back and forth, blurring the distinction between objective and subjective decision-making.

To this end, we take advantage of a unique database of 1.4 million immigration cases processed by ICE between 2012 and 2016. The dataset includes 19 different algorithm versions used consecutively during this period. The data includes detain and release information and end users' dissent[3] to the risk assessment recommendation on each case, allowing us to observe the reaction of officers and supervisors within and across risk algorithms. This data allows us to measure contextual effects on users' dissents, including changes in the size and structure of the caseloads (the users' risk portfolio). More important, as only two versions of these algorithms were used in 2012 and 2016 respectively, we can observe the effect of the electoral cycle as it alters the rates of dissent with the risk assessment made by the algorithm. We show the existence of a political business cycle[4] in punitive bias, distinct from

Significance Statement

This research examines changes in the punitive bias of ICE officers in the 2012 through 2016 period. We provide evidence that contextual factors enter into the immigration decision process and that officers' preferences explain changes to the Risk Classification Assessment. The analysis draws data from 1.4 million immigration detention cases between 2012 and 2016 received pursuant to the Freedom of Information Act (FOIA).

[1] Research on algorithmic justice is increasingly raising attention to biases that are absorbed and propagated by risk assessment tools. Margaret Hu (2017) defines racial biases that enter in the form of "designing, interpreting, and acting" as an algorithmic Jim Crow. Eckhouse et al. (2019) also consider biases in high and low risk classifications, although they do not provide a decomposition model as we do here. See also Mayson (2018) and Huq (2018) for similar arguments.

[2] There is a significant literature on causal attribution and uncertainty that also connects risk and sentencing biases in the decision of justices (Albonetti, 1991; Huber and Gordon, 2004). A different literature has explored the relationship between bias and punishment (Kahneman and Frederick, 2002). Less research has been done on algorithms and algorithmic editing.

[3] Dissent is the word we refer to in discussing the objections that officers and supervisors raise when reviewing the risk recommendation produced by the algorithm. Chiefly it refers to supervisor overrides of the risk recommendation. Officers and supervisors must input an explanation for dissents onto the RCA input screen.

[4] Since Nordhaus (1975) landmark study on the effect of political business cycles, a broad literature measures whether elections affect policy-making and the economy. There is a significant literature on anti-immigration attitudes heightened by the election cycle, but considerable less on the enforcement side of the equation. An early analysis of immigration enforcement and elections by Shughart et al. (1986) provides evidence of "smoothing" within cycles in response to businesses economic pressures. Finally, since Levitt (2002; 1995) landmark paper on electoral cycles and policing, research has modeled crime enforcement and elections (McCrary, 2002). However, no research that we know explains whether punitive biases among immigration officers change with the electoral cycle.

[1] Corresponding author Robert Koulish. E-mail: rkoulish@umd.edu

the mean punitive intent that would be expected to remain steady within algorithms. We also show that the decision by the officers' union to endorse Trump produces a measurable effect on the punitive bias of officers in the days that followed.

Results provide conclusive evidence of contextual factors that enter into the officers' dissent decision, shaping the punitive intent embedded in the system. We show that dissent is higher for cases that the system classifies as low risk and that dissent is lower for cases that the system classifies as high risk, signaling preferences for tougher immigration outcomes by officers and supervisors. We demonstrate that officers are sensitive to caseload features, with dissent changing with the risk portfolio held by each local processing agency, even though the algorithm produced constant output. Finally, we show that officers alter the rate of dissent as elections approach, which is consistent with the political preferences expressed by the Crane v. Napolitano case as well as the endorsement of Donald Trump by the ICE officers' association in 2016.

The organization of this article is as follows. In the first section, we discuss the policy of algorithm "nudges" that guided the implementation of the risk assessment tool by the Obama administration. The algorithm was inspired by academic work on Prospect Theory (Thaler and Sunstein, 2009; Thaler, 2018), which is theoretically and empirically relevant for our analysis. In the second section, we describe the different algorithms implemented by the Obama administration, highlighting how algorithm editing was motivated by the administration's desire to reduce dissent among ICE officers and supervisors. In the third section, we estimate the punitive intent of detention officers using a decomposition Oaxaca-Blinder model that compares dissent to high risk and low risk custody classifications. We show that the theory and the empirics of the case work together, allowing us to describe the punitive intent by officers as it is affected by caseloads and contextual factors. Results of the model provide conclusive evidence of contextual factors shaping the punitive intent. We conclude in section four with the implications of our analyses for the future use of risk algorithms in immigration enforcement.

1. Of Nudges and Dissenters

The immigration risk system was inspired by a national investigation of immigration detention facilities by Dr. Dora Schriro (2009). Like Schriro, many appointees in the Obama Administration favored risk tools as a technique to nudge reform in government regulation. This risk analysis tool was argued for and sponsored by legal scholar and Obama Advisor Cass Sunstein, a colleague of the behavioral economist Richard H. Thaler (Thaler and Sunstein, 2009). Sunstein's immigration "nudge" was designed to mitigate potentially harsh backlashes to regulatory change by ICE officers, as the administration anticipated the punitive views of those in position to enforce immigration rules. Therefore, to limit punitive biases in immigration enforcement, the administration created a risk system tool that was insensitive to personal biases and would nudge decision-making in a less punitive direction.

By the end of Obama's first term, however, plans to soften the harsh impact of immigration enforcement had gone wildly awry. The Obama Administration had become known for detaining and deporting more immigrants annually than all predecessors combined. We have no way to assess what would have been the rate of deportation without the implementation of the "nudge" policy. However, we do know, as the Crane v. Napolitano decision shows, that the Obama's approach to immigration enforcement was not harsh enough for immigration enforcement units within the Administration. Concerns over seemingly lenient turns like Deferred Action for Childhood Arrivals (DACA), Alternatives To Detention (ATD), and even the risk tool itself, hardened into outright opposition by officers in ICE and CBP. The resistance by ICE officers led to frequent algorithm edits, where policy-makers deleted, added, and re-weighted items in the risk assessment tool with the objective of lowering dissent by officers and supervisors. The data, we will show, bears the marks of the conflict between the administration and its enforcement units.

Immigration Enforcement and the Politics of Risk Assessment. Prospect theory, the source of inspiration for the risk assessment algorithm in immigration enforcement, is a framework to explain decisions under uncertainty.[5] It takes as a point of departure the experimental finding that individuals handle gains and losses differently, subject to frames of reference that may be altered by the flow of information. Survey experiments have shown that individuals often weight losses more heavily than gains, with individuals perceiving a higher cost for losing $5 than the utility they receive for gaining it.

As it refers to immigration, uncertainty about the shadow of the future of risk assessment decisions are a textbook example of prospect theory. Let us consider the following example: an officer arrives at the Phoenix processing center one morning and receives two files to evaluate. The first file describes a non-violent immigrant that the system classified as low-risk for flight and low-risk for security. After reviewing the details of the case, the officer agrees with the system recommendation. The second file, on the other hand, describes the detain/release classification of a violent sex offender. The system indicates that the second individual is a high flight and security risk. The officer agrees with both decisions and moves on. However, let us consider what would have been the officer's response if, on that same morning, the case files had been placed in reverse order, allowing the officer to see the most serious case first. Would it be as likely that s/he would agree with the decision to classify the non-violent detainee as a low flight and low security risk?

Risk averse punitive bias describes the officer's belief that the release of a wrongly classified low-risk undocumented immigrant is worse than the detention of a wrongly classified high-risk immigrant. Spillover from worst case examples, then, "nudge" officers in more punitive directions. While we tend to think that immigration cases should be decided on their merits alone, the interpretation of each feature of the case is affected by the frame of reference that is transferred from other cases. We hope that each case will be evaluated without information transfers from other cases, but we do not expect that to be the case. The same set of casefiles may lead to different assessments if officer reads the violent sex offender first, as the change in frames has affected the relative risks of a wrong decision. Consider a very simple model where dissent is a function of the rank assessment given by the algorithm X_i and an unobserved punitive inclination ρ by an officer: $y_i \sim f(\rho, X_i)$. On any given day, we do not expect the particulars of case i

[5] See Thaler (2018) for an excellent review of the history of "nudges" in behavioral economics. See Kahneman and Fredrick (2002) for analyses of biases in punitive assessments. See Bazerman and Moore (2013) for biases in decision-making.

to be associated with those of another case j, $cov(X_i, X_j) = 0$. Therefore, we do not expect that the rank order assessment of case i by the algorithm will modify in any way the assessment of case j. However, prospect theory indicates that conditional and contextual factors in case i will have an effect on the punitive parameter ρ, which will be updated as a function of the initial punitive intent and some information transfer from case j, $\dot{\rho} \sim f(\rho, X_j)$. That is, the unobserved punitive inclination of the officer is updated by the order in which cases are processed, even if the particular of the cases are unrelated to each other. Therefore, information spillovers are not the result of serial correlation across observations but of changes in risk assessments that remain unobserved. These contextual factors reflect information spillovers from other cases (i.e. the risk portfolio observed by officers), but, just as well they could be explained by personal life events, institutional environments, and political contexts.

Punitive intent as an unobserved but changing parameter. Different individuals may hold different punitive inclinations and we may compare them. The punitive inclinations of ICE detention officers, for example, could be the result of self-selection into the security profession, with individuals that seek to "defend their homeland" being recruited at higher rates than those that "value immigration as a social good." Therefore, individuals with a higher punitive inclination than average, $\rho - \bar{\rho} > 0$, are expected to become more prevalent in ICE.

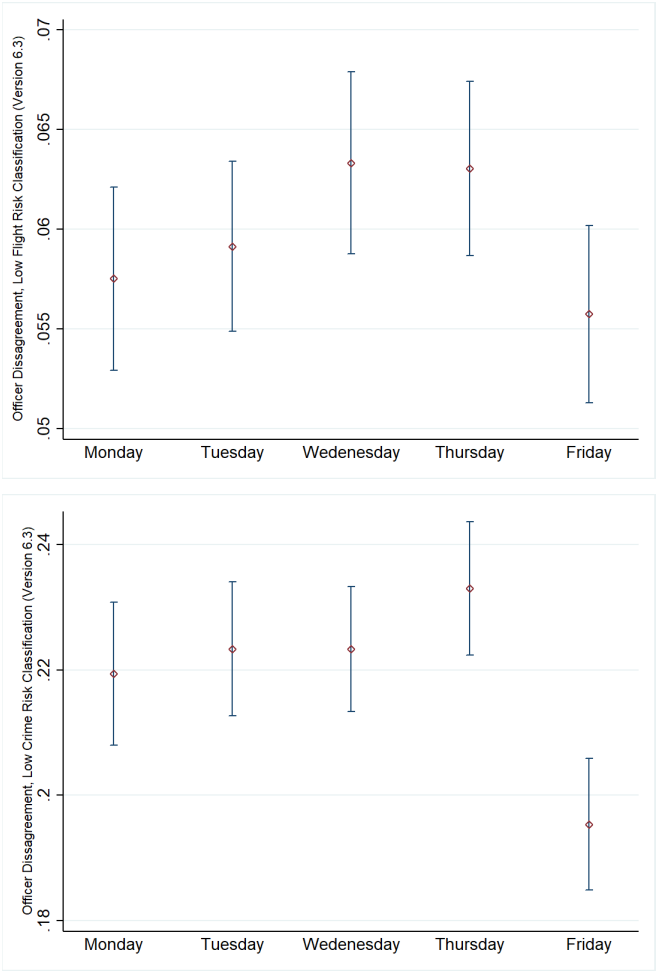
Punitive inclinations may also vary over time, as would be expected under different institutional designs and as a reaction to institutional rewards or sanctions. For example, punitive inclinations may increase if the administration sanctions the failure to prevent the "wrong" type of undocumented immigrant from being released more harshly than the decision to hold the "good" type in custody (Kuisma, 2013). In that case, dissent on low risk assessments may be larger than dissent on high risk assessments, $\rho^{LowRisk} - \rho^{HighRisk} > 0$. In this case, institutional incentives facilitate an update in the average value of ρ because of the fear of public outcry that would ensue.

A nudge, is a device that proposes a punitive baseline $\rho^{baseline}$ and allows for different degrees of deviation from this baseline, making it more costly to act on the punitive biases of the officer. Algorithm editing, on the other hand, is the process by which we modify the punitive baseline to calibrate the nudge. Calibration, however, is difficult as punitive attitudes change in real time, in response to contextual social and political factors such as crimes, crises, or elections.

Nudges may also be observed as political devices and trigger responses that nullify their effect, or worse. While the algorithm hoped to "nudge" punitive inclinations by ICE officers, these goals were rendered visible to officers and explained as a strategy of the current administration. As provision in the system allowed for dissent to be recorded, dissent with the system can become expressive, giving voice to political preferences by the officers that defy the nudge's purpose. Recording the Officers' dissent is clearly great for research, but not so much for policy.

When Nudges get Hungry. "Helium atoms", noted Sabine Hossenfelder (2018), "don't get hungry and are just as well tempered on Monday as they are on Friday". This is certainly not the case with ICE agents and supervisors. The appeal of immigration risk algorithms lies precisely in their capacity to prevent information spillovers that are expected to occur in the detain/release process. While every case should be assessed on its merits, information spillovers result from personal, institutional, social, and political contextual factors that vary over time.

Fig. 1. Lines describe the average dissent rate by ICE officers with the RCA Low Flight Risk (Upper) and Low-Security Risk (Lower) Recommendations, Version 6.3, considering 157,732 cases between May 2015 and October 30, 2016, days before the presidential election. Disagreement is significantly higher for security risk, statistically higher on Thursdays and lower on Fridays.



This is the reason that, different from helium atoms, we expect the officers' rate of dissent with the algorithm recommendations to vary between Monday and Friday, as shown in Figure 1. We also expect dissent rates to be higher when dealing with security decisions compared to flight decisions, as the relative social and professional risks of releasing individuals that may commit a crime differ from those that will simply be lost to the system. Whereas the algorithm recommendation uses fixed weights, our own internal weights adjust to carry-on information, such as prior cases, caseloads, risk levels, the weekly cycles, and the political business cycle.

As the Obama administration introduced its RCA recommendation policy, it recognized that both officers and supervisors would push back, as the new policy constrained

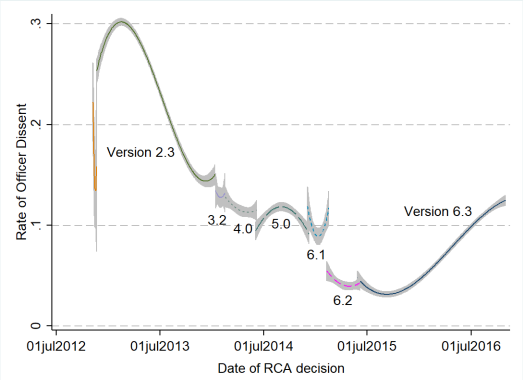


Fig. 2. Overall dissent rates between July 2012 and October 30, 2016, days before the presidential election. Variation within and across algorithms is described in later sections.

their ability to decide on immigration cases. As part of the implementation of the RCA system, therefore, the Obama administration introduced provisions for both officers and supervisors to object. Officers and supervisors were also able to register their written objection to the RCA recommendation and, as important, supervisors were granted authority to make decisions that differed from the RCA recommendation.

Allowing officers and supervisors to dissent with the system recommendation, however, introduced intense pressure to ensure reasonable results. As the system was implemented, high dissent prompted an editing effort by the administration, hoping to ensure that additions, deletions, and re-weighting of risk factors would ensure convergence between the RCA system and the preferences of the officers and the supervisors. As shown in Figure 2, very high rates of dissent in the early days of the system declined over time, reflecting accommodations by both the users and the system.

However, it is worth noticing significant variation over time within each RCA Version, with the rates of dissent shifting even when the algorithm remained unchanged. Within Version 2.3, over the year that follows the 2012 Presidential Election, rates of dissent halved from 30% of officer dissent to a bit over 15%. Similarly, as we moved towards the election of 2016, the rate of dissent among officers tripled, from about 4% to almost 14%.

Some of the variation in dissent rate was expected by a system that hopes to "nudge" people in the right direction. Therefore, we expect dissent to be higher right after implementation. Therefore, it is possible that the significant decline in officer dissent in Version 2.3, after the 2012 election, was due to accommodation between humans and the algorithm. More difficult, however, is to explain the increase in dissent as we approach the 2016 election.

2. Modeling punitive intent using ICE officers' dissent in High and Low Risk cases

We define punitive intent as the difference between the officers' propensity to dissent in low-risk classification cases compared to high-risk classification cases. Our definition borrows from current research that compares high and low record breaking temperatures in global warming models (Meehl et al., 2016). As it will become apparent, the decision to model punitive biases using the difference in the rate of officers' dissent has both clear theoretical and modeling implications, allowing estimation with well-established decomposition models (Oaxaca and Ransom, 1999; O'donnell et al., 2007).

Consider an officer's dissent regression model of the form:

$$y_i \begin{cases} \rho^{High} x_i + e_i^{High} \\ \rho^{Low} x_i + e_i^{Low} \end{cases} \quad [1]$$

where the gap between the different classifications is $y_i^{Low} - y_i^{High} = \rho^{Low} x_i - \rho^{High} x_i$. In Equation (1), y_i describes the observed dissent of an ICE officer; x_i describes a Risk Assessment score given by the algorithm, which summarizes the risk involved in the case according to the administration's priorities. More important, ρ^{Low} and ρ^{High} describe the officers' different tolerance thresholds for cases classified as Low or High risk. Finally, e_i^{Low} and e_i^{High} describe normally distributed error terms.

We conceptualize positive punitive biases as expected dissent rates that are larger for low risk classifications cases compared to high risk cases, $\rho^{Low} - \rho^{High} > 0$. A negative punitive bias, by contrast, describes more frequent dissent in high-risk classifications cases. Therefore, punitive bias is captured by the differences in the unobserved parameters $\rho^{Low} - \rho^{High}$, rather than the overall gap in dissent rates, $y_i^{Low} - y_i^{High}$:

$$y_i \begin{cases} \rho^{Low} - \rho^{High} > 0, \text{ Positive Punitive Bias} \\ \rho^{Low} - \rho^{High} = 0, \text{ Neutral Punitive Bias} \\ \rho^{Low} - \rho^{High} < 0, \text{ Negative Punitive Bias} \end{cases} \quad [2]$$

We may rearrange the terms of Equation (1) as a classic decomposition problem, with the difference in the rate of dissent rate by officers expressed with two separate terms, which capture relative changes in the covariates and tolerance parameters:

$$y_i^{Low} - y_i^{High} = \Delta x \rho^{Low} + \Delta x \rho^{High} \quad [3]$$

Finally, per O'Donnell et. al. (2007), we may use the equivalent twofold formulation that isolates the difference in parameters (punitive bias) and the effect of the covariates (endowments):

$$y_i^{Low} - y_i^{High} = \Delta x \rho^{High} + \Delta \rho x^{Low} \quad [4]$$

The interpretation of the coefficients follows the intuition developed in the previous section. Differences in the rate of dissent by the officers for low and high risk scoring may reflect information associated with the particulars of the case (in-model), including changes in the information entered into the system, caseload, and time constraints, which alter the frequency of the observed explanatory variables. On the other hand, differences in the rate of dissent may reflect unobserved traits in parameters, $\Delta x \rho^{High}$ and the associated changes in attitudes by officers (out-of-model).

As the election approaches, for example, we may observe that the rate of dissent increases for low-risk cases but not for high-risk cases. The change in the rate of dissent, however, may be the result of differences in the covariates, x_i, such as decisions in the type of immigration case that guides arrests, rather than changes in attitudes among ICE officers. Decomposition models allow us to discriminate these competing features of the dissent decision.

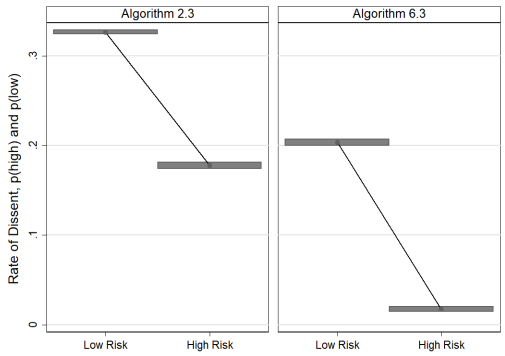


Fig. 3. ICE officers' dissent rate for cases classified by the RCA algorithm as high- or low-security risk. Difference in dissent rates is in Version 2.3 is $y_i^{Low} - y_i^{High} = .3264 - .2032 = .1232$. Difference in Version 6.3 is $y_i^{Low} - y_i^{High} = .1777 - .0172 = .1605$. Therefore, while the overall rate of dissent has declined between versions 2.3 and 6.3, the decline in dissent was much larger when dealing with cases classified as low risk compared to cases classified as high risk.

An Example. Let us exemplify our approach by considering the differences in dissent rates between the Algorithms 2.3 and 6.3, plotted in Figure 3. Readers can see that the average officer dissent in 2016 (Version 6.3) was considerable lower than the one in 2012 (Version 2.3). However, the decline in dissent is much larger for cases classified as high risk compared to cases classified as low risk. That is, while in 2016 almost no officer disagrees with the RCA recommendation for "high safety risk", dissent remained substantive for "low safety risk". The gap in dissent between high and low risk, therefore, increased significantly between Versions 2.3 and 6.3, from .12 to .16, even when the overall dissent rates declined.

Results in Table 1 present the Oaxaca decomposition model with a Gaussian distribution. In the Supplemental Information File (SIF) we present identical results using a logistic distribution. Given that the results do not change, we present results that are more readily interpretable. That will also be the case for all future models, with non-linear logistic versions reported online.

The Oaxaca results in Table 1 provide an intuitive description of the punitive intent as conceptualized in this article, with an increase in the dissent gap explained as the result of on sample features as well as in the punitive bias of ICE officers. Let us describe in detail the results, which will facilitate interpretation of the models that include all covariates.

If we consider the sample of observations in Versions 2.3 and 6.3, the average dissent for low-risk security classification is .292 (29.2%) and the average dissent for high-risk is .0756. Therefore, $y_i^{Low} - y_i^{High} = .216$. The Oaxaca decomposition model distinguishes the effects described in Equation (3), showing that .0715 units of change are explained by sample differences between the two versions, while .149 reflects differences in parameters that are out-of-model. This difference in parameters, $\rho^{Low} - \rho^{High}$, is what we defined as the punitive bias of ICE officers, which can be positive, negative, or equal to zero.

The model in Table 1, of course, lacks most of the covariates of our study. We are not just interested in mean differences between algorithms but, more importantly, in information spillovers and contextual effects.

Table 1. Oaxaca decomposition model with observations from Versions 2.3 and 6.3, as described in the example of Figure 3

	Differential	Covariates (in-Model)	Punitive Bias (out-of-model)
Low-Risk Dissent (Predicted)	0.292*** (0.00131)		
High-Risk Dissent (Predicted)	0.0756*** (0.000824)		
Difference	0.216*** (0.00155)		
RCA Version 6.3		0.0715*** (0.000677)	-0.00394* (0.00163)
Total		0.0715*** (0.000677)	0.145*** (0.00151)
Constant			0.149*** (0.00254)
Observations	223725		

Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

3. Risk Portfolios, Time Constraints, and Contextual Events

Now that we have defined empirically and conceptually the notion of punitive bias, we present the covariates of our study. We consider three families of independent variables: risk portfolio, operational constraints, and contextual events.

We understand changes in risk portfolios as the varying compositions of cases, and their classifications, in a given processing center. As discussed earlier, we expect agents' punitive biases to change with their offices caseload. We explain information spillovers as the result of how the punitive parameter is updated in response to the particulars of recent casefiles, $\dot{\rho} \sim f(\rho, X_j)$. That is, in response to changes in the cumulative load of high-safety risk cases, low-safety risk cases, high-flight risk cases, and low flight risk cases, holding medium risk classification cases as the baseline. As described earlier, we expect officers to adjust their punitive biases when high risk or low risk cases take a larger share of the overall sample.

As the different versions of the algorithm were introduced, and as enforcement priorities changed, the number of cases classified as a low security risk declined sharply (Figure 4). Meanwhile, the number of high-risk cases processed by the system remained relatively constant throughout the entire period. As the risk portfolio of cases changes, we also expect that attention to high and low risk cases will vary.

Different from the effect of risk portfolios, we expect that changes in total caseloads will affect the relative time that officers may spend in each case. Because there are time costs when reading casefiles and writing down dissent, we expect increases in total caseloads to decrease overall dissent and, everything else constant, the punitive bias of the officers. However, given that time constraints reduce investment in reading the particulars of the case, officers' punitive bias may also increase if and when dissent costs are sufficiently low. The tension between cost and bias is well described by Slovic et.al. (2004), showing that the experiential system is fast

Table 2. Correlation matrix, RCA Version 2.3

	Distance to Election 2012	High Risk Flight	Low Risk Flight Low Risk Flight	Case Load High Flight Risk	Case Load Low Flight Risk	Case Load High Security Risk Case Load High Security Risk	Case Load Low Security Risk	Total Number of Cases in Office
Distance to Election 2012	1.000							
High Risk Flight	-0.100	1.000						
	(0.000)							
Low Risk Flight	0.015	-0.279	1.000					
	(0.000)	(0.000)						
Case Load High Flight Risk	-0.314	0.313	-0.093	1.000				
	(0.000)	(0.000)	(0.000)					
Case Load Low Flight Risk	0.029	-0.003	0.019	0.099	1.000			
	(0.000)	(0.273)	(0.000)	(0.000)				
Case Load High Security Risk	-0.008	-0.062	-0.013	0.116	0.252	1.000		
	(0.001)	(0.000)	(0.000)	(0.000)	(0.000)			
Case Load Low Security Risk	-0.312	0.314	-0.092	0.998	0.104	0.063	1.000	
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)		
Total Number of Cases in Office	0.027	0.049	-0.015	0.261	0.038	0.140	0.259	1.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	

Note: P-values in parentheses

Table 3. Correlation matrix, RCA Version 6.3

	Distance to Election 2016	High Risk Flight	Low Risk Flight Low Risk Flight	Case Load High Flight Risk	Case Load Low Flight Risk	Case Load High Security Risk Case Load High Security Risk	Case Load Low Security Risk	Total Number of Cases in Office
Distance to Election 2016	1.000							
High Risk Flight	-0.186	1.000						
	(0.000)							
Low Risk Flight	0.169	-0.837	1.000					
	(0.000)	(0.000)						
Case Load High Flight Risk	-0.392	0.529	-0.447	1.000				
	(0.000)	(0.000)	(0.000)					
Case Load Low Flight Risk	0.151	-0.188	0.169	-0.209	1.000			
	(0.000)	(0.000)	(0.000)	(0.000)				
Case Load High Security Risk	0.075	-0.028	-0.007	0.024	0.637	1.000		
	(0.000)	(0.000)	(0.014)	(0.000)	(0.000)			
Case Load Low Security Risk	-0.409	0.505	-0.420	0.985	-0.219	-0.104	1.000	
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)		
Total Number of Cases in Office	-0.850	0.179	-0.161	0.430	-0.069	-0.019	0.448	1.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	

Note: P-values in parentheses

and mostly automatic, facilitating the expression of bias in decisions (Gigerenzer and Gaissmaier, 2011). Therefore, if the cost of dissent is sufficiently low, more demanding caseloads will reduce analytic assessment in favor of the risk-as-feeling response of the officers. In all, we assume that the cost of dissent remains constant for the entire period and, consequently, assume that tighter time constraints will reduce attention and, therefore, punitive biases.

Fig. 4. Lines describe the average dissent rate by ICE officers with the RCA Low Flight Risk (Upper) and Low-Security Risk (Lower) Recommendations, Version 6.3, considering 157,732 cases between May 2015 and October 30, 2016, days before the presidential election. Disagreement is significantly higher for security risk, statistically higher on Thursdays and lower on Fridays.

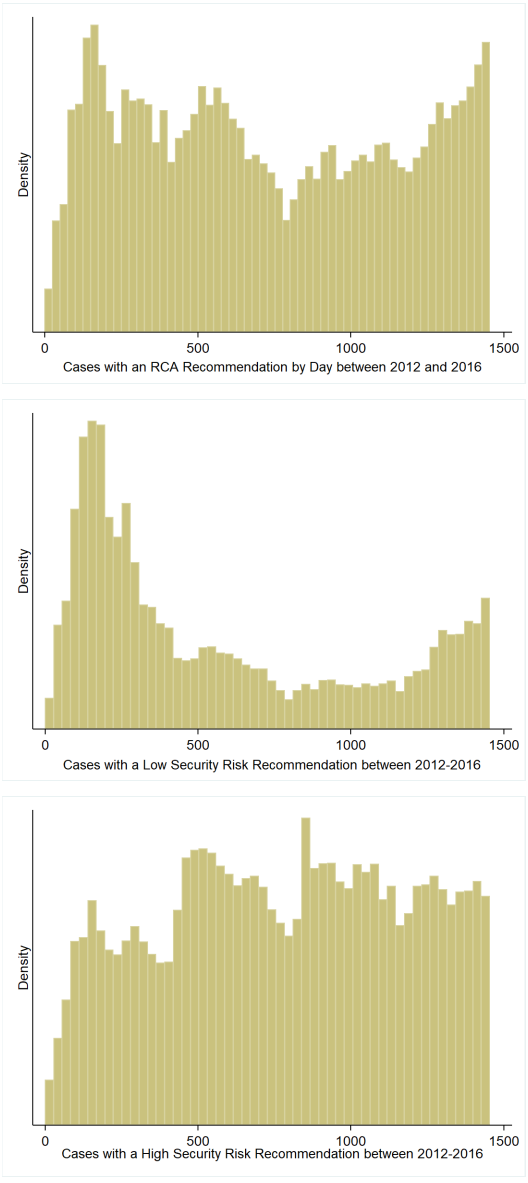


Figure 2 already shows unexpected increases in dissent as we approach the election of 2016. We say unexpected because, different from the Version 2.3 in 2012, the increase in dissent cannot be explained by human accommodation to the algorithm. The "nudge" policy of the Obama administration would be expected to face stronger resistance in the beginning, with dissent higher in the early days of implementation. Officers would first struggle to accommodate the cognitive dissonance prompted by the clash of the algorithm and their own punitive inclinations, with frames adjusting slowly over time. Indeed, accommodation between the algorithm and the officers may be what explains some of the decline in overall dissent across the different versions, from 1.1 through 6.3, beyond the additions, deletions, and re-weighting introduced into the algorithm over time. However, within versions there are no changes in risk assessment. Both the decline in dissent observed in version 2.3 as well as the increase in 6.3 are not due to changes in the risk classification by the system. While the decline in Version 2.3 may be explained away by the officers' adjustment period, the increase in dissent of version 6.3 is unexpected.

A noticeable difference between Versions 2.3 and 6.3, in 2012 and 2016 respectively, is the significant decline in low-low cases (low security and low flight risk). This is clear in the cross-correlation comparison between the covariates in Table 2. While in 2012 the flight risk classifications weakly informed on high-risk classifications, there is a strong positive relationship between high/low flight risk and low/high security risk in 2016. This adjustment, we will show, entered into ICE's officers dissent decision.

The cross-correlations also anticipates why we expect in-model effects in the officers decision that are unrelated to their punitive biases. For example, as the election approaches, there is an increase in the number of high flight risk cases that entered into the system, which have dissent rates that are very different from low flight risk ones. In particular, we can see that a larger number of high flight-low security risk cases entered into the system as we approached the 2016 election.

To distinguish the sample and punitive bias effects of the electoral cycle, we estimate separate decomposition models for Versions 2.3 and 6.3, with a variable that indicates distance in days from the time a case received an RCA recommendation to the closest presidential election. Holding the RCA algorithm constant ensures that the risk classifications do not change within each model. Therefore, non-algorithm features account for all differences within each of the analysis.

Finally, to account for differences in the institutional culture of each processing center, we include fix effects by RCA office. Fixed effects are included in all analyses and reported in the tables, but coefficients were omitted for presentation purposes.

Results. Tables 3 and 4 present results for versions 2.3 and 6.3 of the RCA risk classification tool. As exemplified by Table 1, we direct the attention of readers to three key estimates in each table: (1) the dissent gap, reflected in the difference in the dissent rate for cases classified as low security risk and high security risk; (2) the effect of in-model covariates on the dissent gap; finally, (3) the punitive intent described in the out-of-model column, including the total effect and the contribution by covariates.

Per Equation (3), positive coefficients describe an increase in the dissent gap, explained by in-model covariates as well as out-of-model punitive bias parameters. Our attention is directed to changes in positive out-of-model parameters, that account for the punitive biases of the officers, $\rho^{Low} - \rho^{High} > 0$.

Summary results in Table 3 show that the gap in officer dissent is 14.9 (.149) in Version 2.3, with 32.6% dissent for low security risk classifications and 17.8% for high security risk classifications. As anticipated by Figure 3 earlier in the

Table 4. Oaxaca decomposition model with observations from RCA Version 2.3

	Differential	Covariates (in-model)	Punitive bias (out-of-model)
Low-Risk Dissent (Predicted)	0.326***		
	(0.00159)		
High-Risk Dissent (Predicted)	0.178***		
	(0.00197)		
Difference	0.149***		
	(0.00254)		
Distance to Election 2012		0.00895***	-0.0860***
		(0.000444)	(0.00548)
High Risk Flight		-0.0105***	0.264***
		(0.00118)	(0.00550)
Low Risk Flight		-0.000973***	0.00348***
		(0.000165)	(0.000556)
Case Load High Flight Risk		-0.333***	-0.119**
		(0.0382)	(0.0447)
Case Load Low Flight Risk		0.000191	0.0165***
		(0.000125)	(0.00476)
Case Load High Security Risk		-0.00693**	0.0279
		(0.00231)	(0.0204)
Case Load Low Security Risk		0.454***	0.114**
		(0.0392)	(0.0357)
Total Number of Cases in Office		-0.00353***	-0.100***
		(0.000264)	(0.0135)
Total		0.102***	0.0466***
		(0.00180)	(0.00179)
Constant			-0.138***
			(0.0315)
Observations	124303		

Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 5. Oaxaca decomposition model with observations from RCA Version 6.3

	Differential	Covariates (in-model)	Punitive bias (out-of-model)
Low-Risk Dissent (Predicted)	0.203***		
	(0.00219)		
High-Risk Dissent (Predicted)	0.0172***		
	(0.000508)		
Difference	0.186***		
	(0.00225)		
Distance to Election 2016		0.00873***	-0.0993***
		(0.000708)	(0.00653)
High Risk Flight		0.0200***	-0.167***
		(0.00138)	(0.0266)
Low Risk Flight		-0.0196***	0.0183*
		(0.000832)	(0.00742)
Case Load High Flight Risk		-0.0177	-0.383***
		(0.0380)	(0.0755)
Case Load Low Flight Risk		0.0107***	-0.232***
		(0.00188)	(0.0209)
Case Load High Security Risk		-0.0126***	0.185***
		(0.00339)	(0.0298)
Case Load Low Security Risk		0.0731*	0.242***
		(0.0362)	(0.0638)
Total Number of Cases in Office		-0.00137*	0.125***
		(0.000678)	(0.0273)
Total		0.0870***	0.0990***
		(0.00168)	(0.00168)
Constant			0.414***
			(0.0484)
Observations	99422		

Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

article, Table 3 shows an overall decline in officer dissent in Version 6.3, accompanied by an increase in the gap between high security and low security risk classifications.

We now direct the attention of readers to the in-model column in Table 3, which shows that most of the dissent gap in Version 2.3 is explained by in-model effects of the covariates, 10.2% of the total gap of 14.9%. Meanwhile, changes in punitive intent account for only 4.7% of the total gap.

By contrast, Table 4 shows in Version 6.3 only 8.7% out of a gap of 18.6% is explained in-model, while 9.9% is explained by changes in the punitive bias of the officers. While sample differences in the decomposition models of Table 3 and 4 are not directly comparable, we can see that the out-of-model term explains more of the total gap in the later version of the algorithm.

Let us now consider the effect of within casefile scoring in Version 2.3. We see that punitive biases increase when the algorithm classifies a case as high flight risk, even though flight risk is weakly correlated with high security risk in this version. When the casefile assessed by the officer reports that an undocumented immigrant is a high flight risk, the out-of-model punitive intent increases, $(\rho^{Low} - \rho^{High}|HFR) > \rho^{Low} - \rho^{High}$.

As anticipated in our presentation of the cross-correlations in Table 2, results differ for Version 6.3. There is a positive in-model association between high flight risk and high security risk, which correspond to the changing samples that result from the negative correlation between flight and security in 2016. Given that both classifications are algorithmic, the positive in-model effect of high flight risk in version 6.3 is the result of changes in enforcement priorities and of editing changes between versions. As already reported, examination of the data shows the correlation between High Security Risk and High Flight Risk was -.18 in Version 2.3 and decreased to a much more significant -.61 in version 6.3. That is, flight and security risk become strongly disassociated by 2016. In the later RCA Version, therefore, the out-of-model effect is negative, describing the sample disassociation of high flight and high security risks. As argued earlier in this article, we see the punitive bias of the agent increasing with low flight risk, once it becomes a cue for higher security risk. By contrast, high flight risk become disassociated in the samples and is anticipated by the officers, who reduce their punitive biases.

The portfolio effect for high- and low-flight risk behave as expected. The positive punitive bias in Version 2.3 and the negative punitive bias in Version 6.3 provide evidence of the stronger discrimination in the cases between flight and security risks during the last year of the Obama administration. While officers increase punitive biases in security assessment cases, the out-of-model effect of flight risk dampens punitive biases.

Finally, out-of-model electoral effects are almost identical in 2012 and in 2016. Versions 2.3 and 6.3 show that the further away we are from an election, the lower the punitive bias of ICE officers. Because of changes in the sample of cases being processed, as it was shown in Figure 4, we also observe an in-model increase in the dissent gap as we more away from the election.

The increase in the officers' punitive bias as the election approaches explains two distinct regularities in the data that we could not account for otherwise: first, the increase in dissent for low-risk cases in Version 6.3 is incompatible with an "adjustment" to the algorithm nudge. Rather than the humans becoming accustomed to the classification system, dissent increases over time. Second, this higher dissent rises for low-security risk cases but remains unaltered for high-security risk. In the next section we further test how the political context affects the punitive bias of officers as well as the modifications introduced to the algorithm.

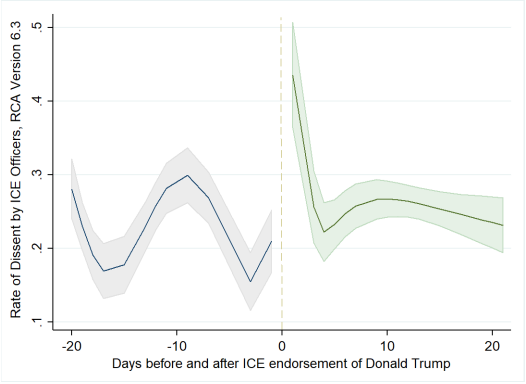


Fig. 5. 5,060 cases received a low-risk for safety score in the 44 days described above and 1,319 received a dissent from the ICE officer in charge (26%). Tuesday after the union endorsed Donald Trump Jr, the rate of dissent increased to 43.5% and on Wednesday to 53.1%.

4. Extensions

In this section, we extend our analysis to explore the effect of political shocks on the punitive gap of ICE officers. We also analyze how dissent leads to changes in the RCA algorithm, which we define as the editing process.

In the first case, we provide evidence of short-term dissent shocks after the union's decision to endorse Donald Trump for president in 2016.[6] In the second case, we provide qualitative evidence to explain the administration's decision to edit the algorithm, which was in response to high rates of dissent by ICE officers. Immigration policy, in this case, was modified to account for the preferences of the enforcers.

The effect of ICE's Decision to Endorse Donald Trump for President. On Monday, September 26 of 2016, the union representing the officers of the Immigration and Customs Enforcement agency publicly announced their support for Donald Trump Jr. to the presidency of the United States. This unprecedented decision surprised many outside the agency, as the level of politicization of representatives and union members was made abundantly clear. It also provides us with an opportunity to examine how contextual events shape the punitive biases by ICE officials.

The decision to endorse Donald Trump was not entirely unexpected. What caught the Council's attention was Trump making immigration restriction a litmus test for his supporters. Trump also reached out for a face-to-face meeting with the Council in which he promised, "to support ICE officers, our nation's laws and our members." The Council turned their back on Hillary Clinton, suggesting her immigration policy

[6] See Di Tella and Schargrodsky (2004) for a natural experiment that exploits data from the aftermath of a terrorist attack to model crime deterrence by the policy. Mastrorocco and Minale (2016) also report results from a natural experiment showing changes in perceptions of crime with media attention. See Weitz-Shapiro and Winters (2017) for an analysis of crime perception based on political preferences and attention.

Table 6. Change in Punitive bias in response to the Union's decision to endorse Donald Trump. Oaxaca Decomposition Model.

	Differential	Explained	Unexplained
Low-Risk Dissent (Predicted)	0.0740*** (0.00252)		
High-Risk Dissent (Predicted)	0.0141*** (0.00211)		
Difference	0.0599*** (0.00329)		
ICE Endorsement		0.00140 (0.00127)	0.0456* (0.0231)
Polynomials (4)		Yes	Yes
Total		-0.00422*** (0.000801)	0.0641*** (0.00339)
Constant			0.0789*** (0.0122)
Observations	13881		

Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

proposals part of a "radical" immigration plan that will lead to the loss of thousands of lives (Politico 9/26/16).

In the 44 days that elapsed between September 4 and October 18 of 2016, 13,881 cases received a Low or High security recommendation, allowing us to model the political shocks in the days that followed the union's decision. Figure 5 provides preliminary evidence that, among the 5,060 cases that received low-risk classification, dissent almost doubled, from an average of 26% to 43.5% on Tuesday and 53.1% on Wednesday after the public announcement.

To analyze the effect of the announcement, we centered the timeline to zero on announcement day and modeled the timelines, before and after, using polynomials (4). We ran the Oaxaca decomposition model using the four polynomials and their interaction with a dummy variable that indicated which RCA decisions occurred after the announcement. Results from the Oaxaca decomposition show an increase in the dissent gap of 5.99%, of which 4.56% is explained out-of-model by the endorsement announcement of ICE. Although the shadow of the future of this union's announcement is short, its fingerprint is unmistakable on the days that followed. Results from the short-term shock of the announcement provide further evidence that the electoral cycle has increased the punitive bias among ICE agents. Over time, however, the shares of risk classifications made by the same algorithm change, as immigration cases that are processed face on-the-ground challenges and as institutional incentives modify enforcement. As immigration enforcement widens, low-risk cases (flight and security) increase in the system. As policy directs attention to high-risk security cases, the share of high-risk flight cases may decline.

Editing the Algorithm: Risk Levels Manipulating the Algorithm. In what follows, we analyze the decisions by the administration to "increase" the punitive bias of the algorithm as a strategy to minimize dissent among ICE officers and supervisors. Quantitative evidence of this shift was described by the decline in dissent rates (Figure 2) as well as the changing ratio of low-risk to high-risk cases processed between 2012 and 2016 (Figure 4). There, we noted a decline in the number of cases classified as low risk from 2012 to 2016. We now describe some of the key policy discussions that informed these adjustments.

Changing enforcement and detention priorities are responsible for much of the manipulation of the algorithm from 2012-2016. During this time, ICE experimented with 19 algorithms in its risk assessment. The algorithm is responsible for categorizing and allocating weight to risk factors. Since many versions represented negligible or no change, for purposes of analysis we condensed the 19 versions to three periods marking substantial shifts in the algorithm: 1) July 2012-January 2014; 2) January 2014-February 2015 3) after February 2015. The changes generally followed the announcement of new ICE enforcement priorities by the ICE Director in a series of memos.

The first stage of the risk assessment led to remarkable reform in the aspirations of ICE detention policy. The stated objective was to land on a consistent and objective technique for assessing detain-and-release decisions. The algorithm allowed for large-scale changes, while ostensibly following what many scholars refer to as objective formal rules. The first stage in the development of the algorithm was bookended by the Schriro Report in 2009 that launched the nudge process followed by a variety of Morton memos designed to implement this new vision for detention, and what was considered an unacceptably high rate of dissent that led to adjustments to lower supervisor overrides in January 2014.

ICE headquarters issued the first batch of Morton memos in 2010 and 2011. The enforcement memoranda were intended to inform and guide the use of ICE resources to arrest, detain, and deport migrants in violation of the immigration laws. ICE then mandated use of the risk tool in detention and release decisions in the form of formal rules, training sessions and FAQ sheets, intended to nudge actual practice in a less punitive direction.

The Memos provided guidance to officers and supervisors in deciding whether to agree or override the RCA recommendations.[7] ICE officers were instructed to prioritize three groups of migrants. The highest priority were individuals "who pose a danger to national security or a risk to public safety." The second level enforcement priorities were individuals who had recently crossed the border illegally or knowingly abused the visa and visa waiver program. The third enforcement category consisted of individuals who had not left the country after a final order of removal, reentered the U.S. illegally after receiving a removal order, or obtained an immigration benefit by fraud.

The nudge was embedded directly in the national RCA training course through references to both the 2011 Prosecutorial Discretion memorandum and the 2011 Enforcement Priorities memorandum (Morton memos). It was also evident in the virtual university and FAQ sheets.

[7] Virtual University, RCA Systems Training, Overview screen 9.

Lessening Dissent. Ulrich Beck (1992) warned that risk science is responsible for making new sets of risks. That seems to be the case during the first stage, as supervisor overrides to the risk recommendation inspired the risk tool architects to modify the algorithm for the purpose of lessening dissent. Thus, the second version of the algorithm largely addressed punitive noise produced by dissent to the initial version. Out of concern for the tension created by dissents, it quieted the noise by accommodating the dissent rather than overcome it. The algorithm accommodated punitive excess in the name of lessening supervisor overrides of the risk recommendation.

In January 2014 ICE headquarters issued a memo to its field offices that announced changes "to strengthen alignment both with ICE priorities and with actual Detain/Release decisions currently being generated by ICE risk officers (ERO Risk end users). The intended outcome was a "decrease in the number of times supervisors need to override RCA recommendations." Among the substantive changes was to "place greater emphasis on aliens' criminal records" so that fewer migrants received a security score of low. This substantial shift in the algorithm accommodating end users is important. It shows that end users could reverse engineer the risk system through dissent. The end users were more punitive in their decisions than in the risk recommendation, producing a higher detention rate for immigrants with criminal records.

Suspending Release. By February 2015, substantial edits to the algorithm reverse engineered the nudge towards the ICE officer's punitive intent. President Obama's November 2014 announcement of Deferred Action for Parents of Arrivals (DAPA), DACA expansion and Felons not Families, a new three-tiered priority enforcement policy, prompted major revisions to the RCA's factors and scoring methodology in February 2015. These revisions were intended to align the RCA with the new prosecutorial priorities.

The shift in Administration priorities got translated at the administrative level into a punitive turn that replaced nearly all flight factors with scoring tied to recency of entry, recency of removal order, and abuse of a visa or visa waiver program. In short, the scoring matrix of February 2015 suspended release for several categories of immigrants, many of whom were clearly no security risk.[8] A guide to the RCA scoring updates in response to President Obama's November 2014 Executive Action announcements makes this realignment clear.[9]

The most significant changes occurred in the public safety component. Many of the factors to measure home stability and community ties, traditional flight risk factors, became largely irrelevant and were replaced with factors keyed to January 1, 2014—the date determined by the Obama Administration as indicating a recent entry or recent order of removal.

By keying the flight risk assessment and to a lesser extent the public safety risk assessment to prosecutorial priorities, the Obama administration made the RCA's detention recommendations reflect its political priorities. This link, however, lacks an underlying logic. Enforcement priorities do not necessarily correlate to public safety or flight risk and thus the need for detention. The RCA, however, no longer distinguished between the two. By eliminating many of the factors assessed that reflect risk and replacing them with policy-based measures, the RCA's algorithm lost the ability to measure true risk, and thus any link to accepted justifications for civil detention. Instead, the RCA became a vehicle to impose detention based on prosecutorial goals (Koulish and Evans, 2019; Evans and Koulish, 2019).

The public safety rubric saw two major changes in the final period of this study. The first was the creation of a new offense severity level: "lowest." The lowest severity level offenses included traffic offenses (except those for hit and run, DUI, and transporting dangerous materials) and a general traffic offense code. These offenses had previously been categorized as low and therefore would have generated 2 points for the public safety score if they were the basis for the ICE encounter, or occurred within the last 5 years prior to the February 2015 changes (See Koulish and Evans (2019)).

This addition finally took offenses like driving without a license—common in states that do not provide drivers licenses to residents without proof of immigration status—out of the public safety risk evaluation. Under the prior rubric, two traffic offenses within the last five years equated to a medium public safety risk level. The February 2015 change recalibrated that assessment to better reflect actual threats to public safety.

The lenient nudge in 2009, however transformed into a punitive shove by 2015. As prospect theory warned, and as contextual factors show, the flow of cases, case levels and the political business cycle ultimately compelled architects of the algorithm to align it with the ICE officers' punitive inclinations, while urging the human factor to consider basing their decisions on such inclinations. In short, the nudge also came with a wink.[10]

Concluding Remarks

How does the punitive bias of enforcement agents change in response to case, office, and contextual events? How does risk assessments tools adapt to the punitive biases of officers? In this article, we provide conclusive evidence that the punitive bias of ICE officers changes in response to contextual factors and that such changes can be expressed in algorithmic editing processes that alter immigration enforcement. In the first part of this article, we provide evidence of changes in punitive biases that affect the processing of immigration cases. In the second part, we describe how pressure by ICE officers informs substantive changes to the different versions of the Risk Classification Algorithm (RCA).

An important methodological contribution of this article is modeling punitive intent as a decomposition problem. Our study describes a theoretical equivalency between punitive biases and the out-of-model effects described by decomposition terms in the Oaxaca-Blinder model. The use of decomposition models to understand implicit biases is, in our view, a strategy generalizable to other risk assessment problems. The out-of-model decomposition terms have a natural interpretation as changes in punitive biases, allowing for a simple integration of theory and empirics. As we show that punitive biases are sensitive to caseloads, risk portfolios, and contextual factors, we contribute to the larger study of risk assessments as a decomposition problem. Model results show that contextual factors shape the punitive intent of officers and have come to guide changes to the detention risk algorithm.

[8] 2016-ICLI-00018 at 15-16 (Feb. 11, 2015) (Email from Marc Rapp, Asst. Dir of Law Enforcement Sys and Analysis to Field Office Dir.'s, Deputy Field Office Dir.'s and Asst. Field Office Dir.'s).

[9] 2016-ICLI-00016 1789 (RCA Executive Action Scoring Updates Guide) (Feb. 2015).

[10] For a sense of the underlying dynamic we are referring to see the "Nudge, Nudge" skit, "A nod (nudge) is as good as a wink to a blind man." Monty Python.

The policy and theoretical implications are equally important. In 2012, Crane et al v. Napolitano became a proxy for opposing Obama and supporting his opponent, Mitt Romney. Romney allies litigated the lawsuit. The litigation targeted key immigration enforcement memos that set Administration priorities behind the risk assessment tool algorithm. The lawsuit alleged that these memos and directives (regarding DACA and the directive to focus on undocumented immigrants with criminal records), had instructed agents to violate the law and Constitution. Eventually, parts of these memos would inform edits introduced to the RCA algorithm, blurring the line between officer's preferences and immigration enforcement.

Named plaintiffs included ICE agents from the ERO office that implements the RCA. The lead litigator, Kris Kobach, who had been ICE Council president, endorsed Romney in January 2012, served as an informal advisor to the Romney campaign, and would also contribute to the GOP's 2012 Republican Platform (Politico, 2/2012; HuffPO8/21/12). The litigation became a litmus test for conservative opposition to Obama immigration policy, as several conservative Republican congress members announced their support for the plaintiffs, including Senator Charles Grassley, and Representative Lamar Smith. Additionally evidence exists that the ICE Council discouraged ICE agents from participating in training on the memos in question (Huffington Post 8/23/12). Few people then guessed that this case along with the ICE officials' union endorsement of Donald Trump would be leading factors in decisions to dissent from the risk algorithm and detain immigrants.

O'Malley (1999) and Zedner (2004) have argued separately, that actuarial risk methods can actually enhance punitive mentalities, rather than stand opposed to them. Feeley and Simon (1992) introduced the term, "new penology," that envisions mass incarceration through risk as a means of managing "aggregates of dangerous groups." Stumpf (2006) has introduced the term crimmigration to demonstrate punitive mentalities at the intersection of criminal law and immigration law. Bosworth and Guild (2008) have further blended punitive mentalities with border criminologies. They envisioned migrants as "a source of potential risks in contemporary political discourse, his/her movements an intelligible object of policing, and his/her body a legitimate object of confinement." (2008). Noferi and Koulish (2014) were first to bring immigration risk into the crimmigration literature. They hypothesized that the detention risk algorithm may legitimize bias and enhance detention. In this article, we provide empirical support to their concerns. Risk algorithms, we show, are malleable and subject to policy whims to accommodate users' subjective judgments. When biased preferences informed edits to the algorithm, this article shows, punitive biases become an enforcement feature.

As much as risk algorithms have the potential to nudge decisions along a softer path, scholars have previously ignored how contextual factors can push back on the nudge, with the tacit approach of the system itself. A message to revel is embedded instruction manuals encouraging officers to consider other factors. Given the institutional makeup of ICE, these words have had the effect of encouraging bias in the face of an objective risk system costing millions of dollars. In the ICE case it undermined the potential for a more humane detention system and undermined the growing dismay over the mistreatment of immigrants in immigration detention.

Going forward it would behoove ICE or any other detention regime to consider the constitutive role that context plays in detention decisions. Risk algorithms are instruments of human beings, and given the highly politicized nature of ICE along with its strong punitive inclinations, a risk algorithm alone is inadequate to the task of lessening mass detention of immigrants.

ACKNOWLEDGMENTS. The authors thank Tiago Ventura, Jose Cabezas, Kate Evans, and Kelsey Drotning for their suggested changes and excellent research support.

References

Albonetti, C. A. (1991). An integration of theories to explain judicial discretion. *Social Problems*, 38(2):247–266.

Bazerman, M. and Moore, D. A. (2013). Judgment in managerial decision making.

Beck, U. (1992). *Risk society: Towards a new modernity*, volume 17. sage.

Bosworth, M. and Guild, M. (2008). Governing through migration control: Security and citizenship in Britain. *The British journal of criminology*, 48(6):703–719.

Di Tella, R. and Schargrodsky, E. (2004). Do police reduce crime? Estimates using the allocation of police forces after a terrorist attack. *American Economic Review*, 94(1):115–133.

Eckhouse, L., Lum, K., Conti-Cook, C., and Ciccolini, J. (2019). Layers of bias: A unified approach for understanding problems with risk assessment. *Criminal Justice and Behavior*, 46(2):185–209.

Evans, K. and Koulish, R. (2019). Manipulating risk in the obama age of immigration detention. *Unpublished Manuscript*.

Feeley, M. M. and Simon, J. (1992). The new penology: Notes on the emerging strategy of corrections and its implications. *Criminology*, 30(4):449–474.

Gigerenzer, G. and Gaissmaier, W. (2011). Heuristic decision making. *Annual review of psychology*, 62:451–482.

Hossenfelder, S. (2018). Lost in math : how beauty leads physics astray.

Hu, M. (2017). Algorithmic Jim Crow. *Fordham L. Rev.*, 86:633.

Huber, G. and Gordon, S. C. (2004). Accountability and coercion: Is justice blind when it runs for office? *American Journal of Political Science*, 48(2):247–263.

Huq, A. Z. (2018). Racial equity in algorithmic criminal justice.

Kahneman, D. and Frederick, S. (2002). Representativeness revisited: Attribute substitution in intuitive judgment. In *Heuristics and biases: The psychology of intuitive judgment.*, pages 49–81. Cambridge University Press, New York, NY, US.

Koulish, R. and Evans, K. (2019). Under Trump, ICE no longer recommends release for immigrants in detentions.

Kuisma, M. (2013). "good" and "bad" immigrants: The economic nationalism of the true finns' immigration discourse. In *The Discourses and Politics of Migration in Europe*, pages 93–108. Springer.

Levitt, S. D. (1995). Using Electoral Cycles in Police Hiring to Estimate the Effect of Policeon Crime. Technical report.

Levitt, S. D. (2002). Using electoral cycles in police hiring to estimate the effects of police on crime: Reply. *American Economic Review*, 92(4):1244–1250.

Mastrorocco, N. and Minale, L. (2016). Information and crime perceptions: evidence from a natural experiment. *Centre for Research and Analysis of Migration (CReAM), Department of Economics, University College London*.

Mayson, S. G. (2018). Bias in, bias out. *Bias Out (September 28, 2018)*, 128.

McCrary, J. (2002). Using electoral cycles in police hiring to estimate the effect of police on crime: Comment. *American Economic Review*, 92(4):1236–1243.

Meehl, G. A., Tebaldi, C., and Adams-Smith, D. (2016). US daily temperature records past, present, and future. *Proceedings of the National Academy of Sciences*, 113(49):13977–13982.

Noferi, M. and Koulish, R. (2014). The immigration detention risk assessment. *Geo. Immigr. LJ*, 29:45.

Nordhaus, W. D. (1975). The political business cycle. *The review of economic studies*, 42(2):169–190.

Oaxaca, R. L. and Ransom, M. R. (1999). Identification in detailed wage decompositions. *Review of Economics and Statistics*, 81(1):154–157.

O'donnell, O., Van Doorslaer, E., Wagstaff, A., and Lindelow, M. (2007). *Analyzing health equity using household survey data: a guide to techniques and their implementation*. The World Bank.

O'Malley, P. (1999). Volatile and contradictory punishment. *Theoretical criminology*, 3(2):175–196.

Schriro, D. B. (2009). *Immigration detention overview and recommendations*. US Department of Homeland Security, Immigration and Customs Enforcement ....

Shughart, W. F., Tollison, R. D., and Kimenyi, M. S. (1986). The political economy of immigration restrictions. *Yale J. on Reg.*, 4:79.

Slovic, P., Finucane, M. L., Peters, E., and MacGregor, D. G. (2004). Risk as analysis and risk as feelings: Some thoughts about affect, reason, risk, and rationality. *Risk analysis*, 24(2):311–322.

Stumpf, J. (2006). The the crimmigration crisis: Immigrants, crime, and sovereign power. *Am. UL Rev.*, 56:367.

Thaler, R. H. (2018). From Cashews to Nudges: The Evolution of Behavioral Economics. *American Economic Review*, 108(6):1265–1287.

Thaler, R. H. and Sunstein, C. R. (2009). *Nudge : improving decisions about health, wealth, and happiness*. Penguin, New York.

Weitz-Shapiro, R. and Winters, M. S. (2017). Can Citizens Discern? Information Credibility, Political Sophistication, and the Punishment of Corruption in Brazil. *The Journal of Politics*, 79(1):60–74.

Zedner, L. (2004). *Criminal justice*. Oxford University Press, Oxford; New York.