Cold Hand Decision-Making *

Daniel Chen
ETH Zurich
Center for Law and Economics

Tobias J. Moskowitz
University of Chicago
Booth School of Business

Kelly Shue
University of Chicago
Booth School of Business

February 21, 2014

Abstract

We document negative autocorrelation in judgment and decision-making in three field contexts: refugee judges, baseball umpires, and loan officers. Previous research on the hot hand fallacy suggests that agents often think of sequential streaks of 0's or 1's as evidence of bias even though such streaks are likely to occur through flips of a fair coin. We hypothesize that the hot hand fallacy and active response to fairness perceptions lead agents to engage in cold hand, i.e. negatively autocorrelated, decision-making. A judge worried about becoming or appearing too lenient or tough if she issues a streak of affirmative or negative decisions may actively adjust her decisions in the opposite direction. We find evidence of negative autocorrelation, particularly among moderate decision-makers, and stronger effects after recent streaks of decisions and in situations when agents face weak incentives for accuracy. We show that our findings are unlikely to be driven by sequential contrast effects and we build a model suggesting the degree of coarse-thinking agents must engage in to explain our results.

Preliminary and Incomplete: Do Not Cite

*We thank Alex Bennett, Kaushik Vasudevan, Chattrin Laksanabunsong, Sarah Eichmeyer, and Luca Braghieri for excellent research assistance and Sue Long for helpful discussions about the asylum court data.

1 Introduction

Decision-makers often operate under uncertainty about what is fair and just. In particular, a judge may be worried that she is becoming or appearing too lenient or tough if she issues a streak of yes or no decisions. We hypothesize that the hot hand fallacy and active response to fairness perceptions lead agents to engage in cold hand, i.e. negatively autocorrelated, decision-making. Empirical research has well documented the hot-hand fallacy, i.e. the belief that a person who has experienced success with a random event will have a greater chance of success in subsequent events (Gilovich et al. 1985; Malkiel 1995) as well as the gambler's fallacy, i.e. the belief that a certain outcome is "due" after a long streak of another outcome (Tversky and Kahneman 1974). These behavioral biases are typically attributed to over-inference from recent streaks of 0's or 1's, even though streaks are likely to occur by chance (Rabin 2002). Agents believe that randomly ordered sequences should have a high rate of alternation while streaks of 0's or 1's are evidence of unfairness or underlying skill/bias.

In this paper, we present evidence that misperception of randomness lead to "cold-handed" decision-making. If cases are ordered randomly, a judge's decision on the previous case should not predict her decision on the next case if decisions are made based upon case merits, controlling for the underlying rate of approval. However, a judge worried that she is becoming or appearing too lenient or tough if she issues a streak of yes or no decisions may engage in negatively autocorrelated decision-making, and fairness considerations may perversely lead to unfair outcomes.

We document cold-hand decision-making in three unique high-stakes settings: the universe of administrative data on refugee court asylum decisions in the US (presented here for the first time), umpire calls in baseball games, and data from a field experiment by Cole et al. (forthcoming) in which Indian loan officers review loan applications. We use these three datasets to show that cold hand bias occurs in a wide variety of contexts and also because each setting offers unique benefits in terms of data analysis. Asylum court decisions offer high frequency decisions that can be linked to judge characteristics. However, the data lacks detailed information about applicant quality, so we cannot assess exactly which decisions represent mistakes. In the baseball data, we have the exact pitch location, so we can control for quality and detect whether the pitch quality itself is negatively autocorrelated. In the loan officers field experiment, we have random assignment of quality of loan applications, and can test how decision-making varies across different monetary incentive schemes. In all three settings, we find evidence of negative autocorrelation in decision making. We estimate that up to 5 percent of decisions are reversed due to cold hand bias. Cold hand bias is significantly stronger among moderate decision-makers who may be ambiguity averse, after recent streaks of past decisions, and in situations in which agents face weak incentives for accuracy. We also link biographical data to decision-makers and corroborate prior research that suggests inexperience magnifies cognitive biases (Krosnick and Kinder 1990; Chen and Berdejó 2013). Finally, we show that our findings of negatively autocorrelated decisions are unlikely to be explained by an alternative behavioral bias: sequential contrast effects. We also conduct tests to show that our results are not driven by non-random variation in underlying case quality. In other words, decisions are negatively autocorrelated but underlying case quality is not.

Our research contributes to the literature on perceptions of fairness. Perceptions of fairness (Rabin 1993) have been incorporated into theoretical and experimental economics, although field data has been relatively scarce (for important exceptions and more general discussion, see Mas (2006); Levitt and List (2007); and Fehr et al. (2009)). Recent theoretical and methodological advances suggest that fairness norms are driven by both social audience and self-image concerns. Andreoni and Bernheim (2009) model the desire to be perceived as fair in the utility function, which gives rise to a signaling game wherein the dictator's choice affects others' inferences about his taste for fairness. Investigation of these issues in the field is challenging due to the fact that methodological designs rely on varying the probability that nature intervenes with a decision.[1] Our research question differs from these studies because we focus on fairness perceptions that perversely lead to unfair or mistaken decision-making. In contrast with previous lab studies, we explore high-stakes decision-making in real-world or field settings. Our research also contributes the sizable psychology literature using vignette studies with small samples of judges that suggest unconscious heuristics (e.g., anchoring, status quo bias, availability) can play a large role in judicial

[1] Andreoni and Bernheim (2009)conduct an experiment where nature sometimes intervenes, choosing an unfavorable outcome for the recipient, and the recipient cannot observe whether nature intervened. They document that subjects are more likely to choose the unfair split, the higher the likelihood that nature intervenes, and interpret fairness behavior as being driven, in part, to subjects' desire to appear fair to others. Chen and Schonger (2013) conduct an experiment where nature sometimes intervenes and makes the decision-maker's choice non-consequential in the strong sense of being not implemented. They document that subjects are more likely to choose the fair split, the higher the likelihood that nature intervenes. Investigation of perceptions of fairness or duty-based motivations in the field is challenging due to the fact that these experimental designs rely on varying the probability that nature intervenes with a decision.

decision-making (see, e.g., Guthrie et al. 2000). To the best of our knowledge, no prior study has used high-frequency judicial decision-making to study whether the kind of phenomenon we describe carries out into actual behavior.

2 Model

2.1 Rational agent

Consider a judge whose job is to decree whether to grant or deny, approve or reject, or call strike (in) or ball (out). We compare what a rational agent would do relative to someone who engages in coarse thinking (Rabin 2002) when evaluating a series of balls that may be insider or outside a boundary. Suppose the balls are generated by a random process $\{y_t\}_{t=1}^{M}$ that consists of a sequence of i.i.d. Bernoulli(α) trials, where α is the probability that the ball falls within the boundaries; in other words, suppose we have a finite sequence of draws $\{y_t\}_{t=1}^{M}$ where $y_t = \{0, 1\}$ (with 1 corresponding to "in" and 0 corresponding to "out"), $P(y_t = 1) = \alpha \in (0, 1)$ and $y_t \perp y_{t+1} \forall t \in \{1, ..., M\}, \ M \in \mathbb{N}$.

The rational agent understands that the y_t are i.i.d. and behaves accordingly.

2.2 Coarse-thinker

There are different degrees of coarse-thinking indexed by $N \in \mathbb{N}, \ N \geq 6$;[2] as will be clear during the course of the model, the lower N the higher the level of coarse thinking.

For rounds $t = 3m - 2, \ m \in \mathbb{N}$, the coarse-thinker believes that balls that she perceives as uncertain are drawn from an urn containing αN "in" signals and $(1 - \alpha) N$ "out" signals.

For rounds $t = 3m - 1, \ m \in \mathbb{N}$, the coarse-thinker believes that balls that she perceives as uncertain are drawn from an urn containing $\alpha N - y_{t-1}$ "in" signals and $(1 - \alpha) N - (1 - y_{t-1})$ "out" signals.

For rounds $t = 3m, \ m \in \mathbb{N}$, the coarse-thinker believes that balls that she perceives as uncertain are drawn from an urn containing $\alpha N - y_{t-1} - y_{t-2}$ "in" signals and $(1 - \alpha) N - (1 - y_{t-1}) - (1 - y_{t-2})$ "out" signals.

The notation is identical to that of Rabin (2002).

[2] $N \geq 6$ because after two draws, the agent must still have a non-zero probability of believing either signal will be drawn next.

3 Predictions

3.1 The coarse-thinker's decisions show a higher degree of negative autocorrelation than the rational thinker's decisions

Lemma 1 Suppose a coarse-thinker is given an sequence of balls $\{y_t\}_{t=1}^{3M+2}$, $M \in \mathbb{N}$, and has to decree whether each ball is "in" or "out". A coarse-thinker believes that the probability that $y_t \neq y_{t+1}$ is greater than the same probability calculated by a rational agent; i.e. $P_c(y_t \neq y_{t+1}) > P_r(y_t \neq y_{t+1})$ where c indexes the coarse-thinker and r indexes the rational agent.

Proof

Rational agent The rational agent understands that $y_t \perp y_{t+1} \forall t \in \{1, ..., 3M+2\}$

$$P_r(y_{t+1} \neq y_t) = \sum_{i=0}^{1} P(y_{t+1} \neq y_t | y_t = i) P(y_t = i) =$$

$$= \sum_{i=0}^{1} P(y_{t+1} = 1 - i) P(y_t = i) =$$

$$= P(y_{t+1} = 1) P(y_t = 0) + P(y_{t+1} = 0) P(y_t = 1) =$$

$$= \alpha(1-\alpha) + (1-\alpha)\alpha = 2\alpha(1-\alpha)$$

Therefore, $P_r(y_{t+1} \neq y_t) = 2\alpha(1-\alpha)$

Coarse-thinker Disregarding the first and last element in the sequence, we can divide the sequence $\{y_t\}_{t=1}^{3M+2}$ in chunks with 3 observations each (i.e. $(y_1, y_2, y_3), (y_4, y_5, y_6)$ etc.); the coarse thinker believes the various triplets are statistically independent. Then, if we take a y_t in the sequence at random while excluding the first and last element, the y_t is equally likely to be in the first position of the triplet, in the second position of the triplet or in the third position. Therefore, all the y_t that have both a predecessor and a successor fall in three mutually exclusive and exhaustive cases, namely $t \in \{n | n = 3m-1, m \in \mathbb{N}\}$, $t \in \{n | n = 3m-2, m \in \mathbb{N}\}$ and $t \in \{n | n = 3m, m \in \mathbb{N}\}$, with

$$P(t \in \{n | n = 3m - i, m \in \mathbb{N}\}) = \frac{1}{3}, \ i = \{0, 1, 2\}$$

In order to find $P_c\left(y_t \neq y_{t+1}\right)$, we condition on the position of t within the triplet: first position in the triplet, second position in the triplet or third position in the triplet

$$P_c\left(y_{t+1} \neq y_t\right) = \sum_{i=0}^{2} P\left(y_{t+1} \neq y_t | t \in \{n | n = 3m - i, m \in \mathbb{N}\}\right) P\left(t \in \{n | n = 3m - i, m \in \mathbb{N}\}\right) =$$

$$= \frac{1}{3} \sum_{i=0}^{2} P\left(y_{t+1} \neq y_t | t \in \{n | n = 3m - i, m \in \mathbb{N}\}\right)$$

Let's start by considering the first position in the triplet; i.e. let's start with $P\left(y_{t+1} \neq y_t | t \in \{n | n = 3m - 2, m \in\right.$

$$P\left(y_{t+1} \neq y_t | t \in \{n | n = 3m - 2, m \in \mathbb{N}\}\right) =$$

$$= \sum_{i=0}^{1} P\left(y_{t+1} \neq y_t | y_t = i, t \in \{n | n = 3m - 2, m \in \mathbb{N}\}\right) P\left(y_t = i | t \in \{n | n = 3m - 2, m \in \mathbb{N}\}\right) =$$

$$= \sum_{i=0}^{1} P\left(y_{t+1} = 1 - i | y_t = i, t \in \{n | n = 3m - 2, m \in \mathbb{N}\}\right) P\left(y_t = i | t \in \{n | n = 3m - 2, m \in \mathbb{N}\}\right) =$$

$$= (1 - \alpha) \frac{\alpha N}{N - 1} + \alpha \frac{(1 - \alpha) N}{N - 1} = 2\alpha (1 - \alpha) \frac{N}{N - 1}$$

Let's now consider the second position in the triplet; i.e. let's consider $P\left(y_{t+1} \neq y_t | t \in \{n | n = 3m - 1, m \in \mathbb{N}\}\right)$:

$$P\left(y_{t+1} \neq y_t | t \in \{n | n = 3m - 1, m \in \mathbb{N}\}\right) =$$

$$= \sum_{i=0}^{1} \sum_{j=0}^{1} P\left(y_{t+1} = 1 - i | y_{t-1} = j, y_t = i, t \in \{n | n = 3m - 1, m \in \mathbb{N}\}\right) P\left(y_{t-1} = j, y_t = i | t \in \{n | n = 3m - 1, m \in \mathbb{N}\right.$$

$$= (1 - \alpha) \frac{(1 - \alpha) N - 1}{N - 1} \frac{\alpha N}{N - 2} + \alpha \frac{(1 - \alpha) N}{N - 1} \frac{\alpha N - 1}{N - 2} + (1 - \alpha) \frac{\alpha N}{N - 1} \frac{(1 - \alpha) N - 1}{N - 2} + \alpha \frac{\alpha N - 1}{N - 1} \frac{(1 - \alpha) N}{N - 2} =$$

$$= 2 (1 - \alpha) \frac{(1 - \alpha) N - 1}{N - 1} \frac{\alpha N}{N - 2} + 2\alpha \frac{(1 - \alpha) N}{N - 1} \frac{\alpha N - 1}{N - 2} =$$

$$= \frac{2\alpha (1 - \alpha) N}{(N - 1)(N - 2)} \left[(1 - \alpha) N - 1 + \alpha N - 1\right] =$$

$$= 2\alpha (1 - \alpha) \frac{N}{N - 1}$$

Let's finally consider the third position in the triplet; i.e. let's consider $P\left(y_{t+1} \neq y_t | t \in \{n | n = 3m-1, m \in \mathbb{N}\}\right)$:

$$P\left(y_{t+1} \neq y_t | t \in \{n | n = 3m, m \in \mathbb{N}\}\right) =$$

$$= \sum_{i=0}^{1}\sum_{j=0}^{1}\sum_{k=0}^{1}\left[P\left(y_{t+1} = 1-i | y_{t-2} = k, y_{t-1} = j, y_t = i, t \in \{n | n = 3m, m \in \mathbb{N}\}\right) \times\right.$$

$$\left. P\left(y_{t-2} = k, y_{t-1} = j, y_t = i | t \in \{n | n = 3m, m \in \mathbb{N}\}\right)\right] =$$

$$= \sum_{i=0}^{1}\sum_{j=0}^{1}\sum_{k=0}^{1}\left[P\left(y_{t+1} = 1-i | t \in \{n | n = 3m, m \in \mathbb{N}\}\right) P\left(y_{t-2} = k, y_{t-1} = j, y_t = i | t \in \{n | n = 3m, m \in \mathbb{N}\}\right)\right] =$$

$$= \alpha\left[(1-\alpha)\frac{(1-\alpha)N-1}{N-1}\frac{(1-\alpha)N-2}{N-2} + \alpha\frac{(1-\alpha)N}{N-1}\frac{(1-\alpha)N-1}{N-2} + \right.$$

$$\left. + (1-\alpha)\frac{\alpha N}{N-1}\frac{(1-\alpha)N-1}{N-2} + \alpha\frac{\alpha N-1}{N-1}\frac{(1-\alpha)N}{N-2}\right] +$$

$$+ (1-\alpha)\left[(1-\alpha)\frac{(1-\alpha)N-1}{N-1}\frac{\alpha N}{N-2} + \alpha\frac{(1-\alpha)N}{N-1}\frac{\alpha N-1}{N-2} + (1-\alpha)\frac{\alpha N}{N-1}\frac{\alpha N-1}{N-2} + \alpha\frac{\alpha N-1}{N-1}\frac{\alpha N-2}{N-2}\right] =$$

$$= 2\alpha(1-\alpha)$$

Therefore,

$$P_c\left(y_t \neq y_{t+1}\right) = \frac{1}{3}\left[2\alpha(1-\alpha)\frac{N}{N-1} + 2\alpha(1-\alpha)\frac{N}{N-1} + 2\alpha(1-\alpha)\right] =$$

$$= 2\alpha(1-\alpha)\frac{3N-1}{3(N-1)}$$

Finally, let's check whether $P_c\left(y_t \neq y_{t+1}\right) > P_r\left(y_t \neq y_{t+1}\right)$

$$P_c\left(y_t \neq y_{t+1}\right) > P_r\left(y_t \neq y_{t+1}\right)$$

$$2\alpha(1-\alpha)\frac{3N-1}{3(N-1)} > 2\alpha(1-\alpha)$$

$$3N-1 > 3(N-1)$$

$$2 > 0$$

which evaluates to True. Therefore, $P_c\left(y_t \neq y_{t+1}\right) > P_r\left(y_t \neq y_{t+1}\right)$.$\square$

Discussion Note that the proof does not rely on agents being equally likely to be in a period where the coarse-thinker knows the urn is being renewed, or, where the coarse-thinker knows the urn has had one or two draws already.

3.2 Moderate decision makers show stronger negative autocorrelation than extreme decision makers

Define a moderate decision maker as the umpire for whom $\alpha = \frac{1}{2}$. The intuition is that an umpire is considered moderate if, whenever she perceives the ball as uncertain, she decrees the ball is "in" $\frac{1}{2}$ of the times and "out" the other $\frac{1}{2}$ of the times.

Lemma 2 Suppose a coarse-thinker is given an sequence of balls $\{y_t\}_{t=1}^{3M+2}$, $M \in \mathbb{N}$, and has to decree whether each ball is "in" or "out". Then, for a coarse-thinker,

$$\arg\max_{\alpha\in(0,1)} P_c\left(y_t \neq y_{t+1}\right) = \frac{1}{2}$$

i.e. moderate coarse thinkers have the highest probability of decreeing $y_t \neq y_{t+1}$ among the coarse thinkers.

Proof This is a simple maximization problem. From Lemma 1, we know that

$$P_c\left(y_t \neq y_{t+1}\right) = 2\alpha\left(1-\alpha\right)\frac{3N-1}{3\left(N-1\right)}$$

The FOC with respect to α is

$$\frac{dP_c\left(y_t \neq y_{t+1}\right)}{d\alpha} = 0 \iff \frac{2}{3}\frac{3N-1}{N-1}\left(1-2\alpha\right) = 0$$

$$\alpha = \frac{1}{2}$$

The SOC with respect to α is

$$-\frac{4}{3}\frac{3N-1}{N-1} < 0$$

which is always true because $N \geq 6$.

Therefore,

$$\underset{\alpha\in(0,1)}{\arg\max} P_c\left(y_t \neq y_{t+1}\right) = \frac{1}{2}$$

□

3.3 The coarse-thinker's decisions show a higher degree of negative autocorrelation after streaks of decisions in the same direction

Lemma 3 Suppose a coarse-thinker is given an sequence of balls $\{y_t\}_{t=1}^{3M+2}$, $M \in \mathbb{N}$, and has to decree whether each ball is "in" or "out"; then, $\exists\left(\underline{\alpha}, \bar{\alpha}\right)$ *s.t.* $\forall \alpha \in \left(\underline{\alpha}, \bar{\alpha}\right)$ *and* $\forall N \geq 6$

$$\frac{P_c\left(y_t \neq y_{t+1} | y_t = y_{t-1}\right)}{P_r\left(y_t \neq y_{t+1} | y_t = y_{t-1}\right)} > \frac{P_c\left(y_t \neq y_{t+1}\right)}{P_r\left(y_t \neq y_{t+1}\right)}$$

Proof

Rational agent

$$P_r\left(y_t \neq y_{t+1} | y_t = y_{t-1}\right) = \frac{P\left(y_t \neq y_{t+1} \wedge y_t = y_{t-1}\right)}{P\left(y_t = y_{t-1}\right)}$$

Since the outcome is binary, there are two possibilities: either y_{t-1} and y_t equal 1 and y_{t+1} equals 0 or viceversa. Therefore,

$$P_r\left(y_t \neq y_{t+1} | y_t = y_{t-1}\right) = \frac{P\left(y_{t-1} = 0, y_t = 0, y_{t+1} = 1\right) + P\left(y_{t-1} = 1, y_t = 1, y_{t+1} = 0\right)}{P\left(y_t = 0, y_{t-1} = 0\right) + P\left(y_t = 1, y_{t-1} = 1\right)} =$$

$$= \frac{(1-\alpha)^2 \alpha + \alpha^2 (1-\alpha)}{(1-\alpha)^2 + \alpha^2} = \frac{\alpha (1-\alpha)}{(1-\alpha)^2 + \alpha^2}$$

Coarse-thinker Disregarding the first and last element in the sequence, we can divide the sequence $\{y_t\}_{t=1}^{3M+2}$ in chunks with 3 observations each (i.e. $\left(y_1, y_2, y_3\right), \left(y_4, y_5, y_6\right)$ etc.); the coarse thinker believes the various triplets are statistically independent. Then, if we take a y_t in the sequence at random while excluding the first and last element, the y_t is equally likely to be in the first position of the triplet, in the second position of the triplet or in the third position. Therefore, all the y_t that have both a predecessor and a successor fall in three mutually exclusive and exhaustive cases, namely $t \in \{n | n = 3m-1, m \in \mathbb{N}\}$, $t \in \{n | n = 3m-2, m \in \mathbb{N}\}$ and $t \in \{n | n = 3m, m \in \mathbb{N}\}$, with

$$P\left(t\in\{n|n=3m-i,m\in\mathbb{N}\}\right)=\frac{1}{3},\ i=\{0,1,2\}$$

In order to find $P_c\left(y_t\neq y_{t+1}|y_{t-1}=y_t\right)$, we condition on the position of t within the triplet: first position in the triplet, second position in the triplet or third position in the triplet

$$P\left(y_t\neq y_{t+1}|y_{t-1}=y_t\right)=\sum_{i=0}^{2}P\left(y_t\neq y_{t+1}|y_{t-1}=y_t,t\in\{n|n=3m-i,m\in\mathbb{N}\}\right)P\left(t\in\{n|n=3m-i,m\in\mathbb{N}\}\right)=$$

$$=\frac{1}{3}\sum_{i=0}^{2}P\left(y_t\neq y_{t+1}|y_{t-1}=y_t,t\in\{n|n=3m-i,m\in\mathbb{N}\}\right)$$

Let's start by considering the first position in the triplet; i.e. let's start with $P\left(y_t\neq y_{t+1}|y_{t-1}=y_t,t\in\{n|n=3m-\right.$

$$P\left(y_t\neq y_{t+1}|y_{t-1}=y_t,t\in\{n|n=3m-2,m\in\mathbb{N}\}\right)=$$

$$=\sum_{i=0}^{1}\sum_{j=0}^{1}\sum_{k=0}^{1}\left[\frac{P\left(y_{t-3}=k,y_{t-2}=j,y_{t-1}=i,y_t=i,y_{t+1}=1-i|t\in\{n|n=3m-2,m\in\mathbb{N}\}\right)}{P\left(y_{t-1}=i,y_t=i\right)}\right]=$$

$$=\left\{\left[(1-\alpha)\frac{(1-\alpha)N-1}{N-1}\frac{(1-\alpha)N-2}{N-2}+\alpha\frac{(1-\alpha)N}{N-1}\frac{(1-\alpha)N-1}{N-2}+\right.\right.$$

$$\left.+(1-\alpha)\frac{\alpha N}{N-1}\frac{(1-\alpha)N-1}{N-2}+\alpha\frac{\alpha N-1}{N-1}\frac{(1-\alpha)N}{N-2}\right](1-\alpha)\frac{\alpha N}{N-1}+$$

$$+\left[(1-\alpha)\frac{(1-\alpha)N-1}{N-1}\frac{\alpha N}{N-2}+\alpha\frac{(1-\alpha)N}{N-1}\frac{\alpha N-1}{N-2}+\right.$$

$$\left.\left.+(1-\alpha)\frac{\alpha N}{N-1}\frac{\alpha N-1}{N-2}+\alpha\frac{\alpha N-1}{N-1}\frac{\alpha N-2}{N-2}\right]\alpha\frac{(1-\alpha)N}{N-1}\right\}/$$

$$\left\{\left[(1-\alpha)\frac{(1-\alpha)N-1}{N-1}\frac{(1-\alpha)N-2}{N-2}+\alpha\frac{(1-\alpha)N}{N-1}\frac{(1-\alpha)N-1}{N-2}+\right.\right.$$

$$\left.+(1-\alpha)\frac{\alpha N}{N-1}\frac{(1-\alpha)N-1}{N-2}+\alpha\frac{\alpha N-1}{N-1}\frac{(1-\alpha)N}{N-2}\right](1-\alpha)+$$

$$+\left[(1-\alpha)\frac{(1-\alpha)N-1}{N-1}\frac{\alpha N}{N-2}+\alpha\frac{(1-\alpha)N}{N-1}\frac{\alpha N-1}{N-2}+\right.$$

$$\left.\left.+(1-\alpha)\frac{\alpha N}{N-1}\frac{\alpha N-1}{N-2}+\alpha\frac{\alpha N-1}{N-1}\frac{\alpha N-2}{N-2}\right]\alpha\right\}=$$

$$=\frac{N}{N-1}\frac{\alpha(1-\alpha)}{(1-\alpha)^2+\alpha^2}$$

Let's move to the second position in the triplet; i.e. let's consider $P\left(y_t \neq y_{t+1} | y_{t-1} = y_t, t \in \{n | n = 3m-1, m \in \mathbb{N}\}\right.$

$$P\left(y_t \neq y_{t+1} | y_{t-1} = y_t, t \in \{n | n = 3m-1, m \in \mathbb{N}\}\right) =$$

$$= \sum_{i=0}^{1} \left[\frac{P\left(y_{t-1} = i, y_t = i, y_{t+1} = 1-i | t \in \{n | n = 3m-1, m \in \mathbb{N}\}\right)}{P\left(y_{t-1} = i, y_t = i\right)} \right] =$$

$$= \frac{(1-\alpha) \frac{(1-\alpha)N-1}{N-1} \frac{\alpha N}{N-2} + \alpha \frac{\alpha N-1}{N-1} \frac{(1-\alpha)N}{N-2}}{(1-\alpha) \frac{(1-\alpha)N-1}{N-1} + \alpha \frac{\alpha N-1}{N-1}} =$$

$$= \frac{\alpha (1-\alpha) N}{(N-1) - 2\alpha (1-\alpha) N}$$

Let's finally consider the third position in the triplet; i.e. let's consider $P\left(y_t \neq y_{t+1} | y_{t-1} = y_t, t \in \{n | n = 3m, m \in \mathbb{N}\right.$

$$P\left(y_t \neq y_{t+1} | y_{t-1} = y_t, t \in \{n | n = 3m-1, m \in \mathbb{N}\}\right) =$$

$$= \sum_{i=0}^{1} \sum_{j=0}^{1} \left[\frac{P\left(y_{t-2} = j, y_{t-1} = i, y_t = i, y_{t+1} = 1-i | t \in \{n | n = 3m-2, m \in \mathbb{N}\}\right)}{P\left(y_{t-1} = i, y_t = i\right)} \right] =$$

$$= \left\{ \left[(1-\alpha) \frac{(1-\alpha)N-1}{N-1} \frac{(1-\alpha)N-2}{N-2} + \alpha \frac{(1-\alpha)N}{N-1} \frac{(1-\alpha)N-1}{N-2} \right] \alpha + \right.$$

$$\left. + \left[(1-\alpha) \frac{\alpha N}{N-1} \frac{\alpha N-1}{N-2} + \alpha \frac{\alpha N-1}{N-1} \frac{\alpha N-2}{N-2} \right] (1-\alpha) \right\} /$$

$$\left[(1-\alpha) \frac{(1-\alpha)N-1}{N-1} \frac{(1-\alpha)N-2}{N-2} + \alpha \frac{(1-\alpha)N}{N-1} \frac{(1-\alpha)N-1}{N-2} + \right.$$

$$\left. + (1-\alpha) \frac{\alpha N}{N-1} \frac{\alpha N-1}{N-2} + \alpha \frac{\alpha N-1}{N-1} \frac{\alpha N-2}{N-2} \right] =$$

$$= \frac{\alpha (1-\alpha)(N-2)}{(N-1) - 2\alpha (1-\alpha) N}$$

Therefore

$$P\left(y_t \neq y_{t+1} | y_{t-1} = y_t\right) = \frac{1}{3} \sum_{i=0}^{2} P\left(y_t \neq y_{t+1} | y_{t-1} = y_t, t \in \{n | n = 3m-i, m \in \mathbb{N}\}\right) =$$

$$= \frac{1}{3} \left[\frac{N}{N-1} \frac{\alpha (1-\alpha)}{(1-\alpha)^2 + \alpha^2} + \frac{\alpha (1-\alpha) N}{(N-1) - 2\alpha (1-\alpha) N} + \frac{\alpha (1-\alpha)(N-2)}{(N-1) - 2\alpha (1-\alpha) N} \right] =$$

$$= \frac{1}{3}\alpha(1-\alpha)\left[\frac{N}{N-1}\frac{1}{(1-\alpha)^2+\alpha^2} + \frac{2(N-1)}{(N-1)-2\alpha(1-\alpha)N}\right]$$

Finally, let's check whether $\exists(\underline{\alpha},\bar{\alpha})\ s.t.\ \forall\alpha \in (\underline{\alpha},\bar{\alpha})\ and\ \forall N \geq 6$,

$$\frac{P_c(y_t \neq y_{t+1}|y_t = y_{t-1})}{P_r(y_t \neq y_{t+1}|y_t = y_{t-1})} > \frac{P_c(y_t \neq y_{t+1})}{P_r(y_t \neq y_{t+1})}$$

$$\frac{\frac{1}{3}\alpha(1-\alpha)\left[\frac{N}{N-1}\frac{1}{(1-\alpha)^2+\alpha^2} + \frac{2(N-1)}{(N-1)-2\alpha(1-\alpha)N}\right]}{\frac{\alpha(1-\alpha)}{(1-\alpha)^2+\alpha^2}} > \frac{2\alpha(1-\alpha)\frac{3N-1}{3(N-1)}}{2\alpha(1-\alpha)}$$

$$\frac{1}{6}\left(3-\sqrt{3}\right) \leq \alpha \leq \frac{1}{6}\left(3+\sqrt{3}\right)$$

Letting $\underline{\alpha} = \frac{1}{6}\left(3-\sqrt{3}\right) \approx 0.21$ and $\bar{\alpha} = \frac{1}{6}\left(3+\sqrt{3}\right) \approx 0.79$, we have shown that $\exists(\underline{\alpha},\bar{\alpha})\ s.t.\ \forall\alpha \in (\underline{\alpha},\bar{\alpha})\ and\ \forall N \geq 6$

$$\frac{P_c(y_t \neq y_{t+1}|y_t = y_{t-1})}{P_r(y_t \neq y_{t+1}|y_t = y_{t-1})} > \frac{P_c(y_t \neq y_{t+1})}{P_r(y_t \neq y_{t+1})}$$

□

Discussion Extreme judges will not do streaks for any level of course thinking. As α approaches 0, the expression reduces to $3N-2 > 3N-1$, an impossibility. When $\alpha = \frac{1}{2}$, then the expression reduces to $N > 0$. That is, for any finite N level of course-thinking, the perfectly moderate judge will react to streaks. Note that the level of α where agents react to streaks indicates the level of N degree coarse-thinking that an agent engages in.

3.4 For high N, the coarse-thinker behaves more like the rational agent

Lemma 4

$$\frac{d\left[\frac{P_c(y_t \neq y_{t+1})}{P_r(y_t \neq y_{t+1})}\right]}{dN} < 0$$

and

$$\lim_{N\to\infty} P_c(y_t \neq y_{t+1}) = P_r(y_t \neq y_{t+1})$$

Proof

$$\frac{d\left[\frac{P_c(y_t \neq y_{t+1})}{P_r(y_t \neq y_{t+1})}\right]}{dN} = \frac{d\left[\frac{2\alpha(1-\alpha)\frac{3N-1}{3(N-1)}}{2\alpha(1-\alpha)}\right]}{dN} =$$

$$= \frac{d\left[\frac{3N-1}{3(N-1)}\right]}{dN} = -\frac{2}{3\left(N-1\right)^2} < 0$$

and

$$\lim_{N\to\infty} P_c\left(y_t \neq y_{t+1}\right) = \lim_{N\to\infty} 2\alpha\left(1-\alpha\right)\frac{3N-1}{3\left(N-1\right)} =$$

$$= 2\alpha\left(1-\alpha\right) = P_r\left(y_t \neq y_{t+1}\right)$$

□

Discussion Experience, education, incentives could all increase N.

3.5 Self-certainty

Another way to model experience and education is that agents may be more likely to believe some balls are certainly in or certainly out. This self-certainty would also tend to attenuate but not reverse the predictions of the model, if these balls are not from the urn drawn with replacement. The coarse thinker is less negatively autocorrelated compared to when the current ball comes after an uncertain ball. The intuition is that there is no information obtained from a certain ball about the subsequent ball.

However, if agents believe that the certain balls are also drawn from the urn without replacement, then the coarse thinker is more negatively autocorrelated than when the previous ball is uncertain. The certain ball now contains more information about the subsequent ball than an uncertain ball. Proof is available on request. Formally, negative autocorrelation after an uncertain ball is generally more muted than the negative autocorrelation after a certain ball even for a rational thinker, so the coarse thinker has less room for bias.

4 Data Description and Institutional Context

4.1 Refugee Courts

We use the universe of administrative data on US refugee asylum cases over 20 years across 50 courthouses. Judges are randomly assigned assigned to cases and decide whether to grant or deny asylum. This data contains the exact time when a decision was made, the identity of the judge, and litigant characteristics. The data includes over 15 million hearing sessions on over 3 million decisions. We know when the asylum case was assigned, whether the hearing was an individual hearing or whether multiple individuals were scheduled in the same session, how many cases were scheduled for sessions during a day for that judge, whether this was an in person hearing or by audio or video, whether it was a written or oral order, whether there are other related applications for relief filed by the individual and the judge's ruling on each, ethnicity of the applicant, the reason for the case, and the judge.

We merge the decision and hearing datasets to determine whether the decision occurred on the date of the last hearing session or after it. We keep decisions whose final hearing session occurred on or before the decision completion date. We excluded non-asylum decisions: that is, we focus on applications for asylum, asylum-withholding, or withholding-convention against torture. When an individual had multiple decisions on the same day on these three applications, we focus on one decision in the order listed above, as asylum decisions would be the most salient. 93% of the resulting data are represented by asylum decisions and most individuals have all applications on the same day denied or granted. We merge this data with judicial biographies that we augmented.

We exclude family members except the lead family member because in almost all cases, all family members are either granted or denied asylum together. Family members are inferred if individuals in the same city shared national origin, had decisions on the same day, had the same grant outcome, had the same indicator for having a lawyer represent them and the same indicator for whether the case was defensive or not (i.e., lodged in defense against a removal proceeding initiated by the government). All other decisions on these family members were flagged.

We also exclude and flag individuals who see the same judge at the same clock time (called master sessions) and individuals who see multiple judges at the same clock time. Any decision that comes after any flagged decision were also excluded because it is not clear what is the lagged decision, except when the decision comes after family members that were seen consecutively. Finally, we restrict our sample to decisions whose immediately prior decision by the judge is on the same day or previous day or over the weekend if it is a Monday decision. The data spans 602,500 decisions from 1971-2013, although coverage is not thick until 1985. Applying all the exclusions except recency results in 279,145 decisions. Recency restricts to 106,071 decisions. Restricting further to decisions made by moderate judges (those whose average grant rate excluding the day of the decision is between 0.3 and 0.7) results in 48,930 decisions.

In general, judges review cases in a first-in-first out fashion. Cases filed within each city are randomly assigned to judges within each city, and judges review cases in the order in which they are assigned. Exceptions to this rule occur when applicants are heard multiple times, file applications on additional issues, get delays, and have closures made that are other than grant or deny (e.g. the applicant doesn't show up, withdraws, and an "other" category that is miscellaneous). We assume that these violations of first-in-first-out, which are likely driven by applicant behaviors, are uncorrelated with the judge's lagged decision.

4.2 Baseball

We use data on umpire calls of pitches from PITCHf/x, a system that tracks the speeds and trajectories of pitched baseballs. The system was installed in 2006 in every Major League Baseball stadium. The system records the path and speed of each pitch, as well as the location with respect to the strike zone as the pitch crossed the front of the home plate. The PITCHf/x data precisely records whether each pitch was within the strike zone. This data can be used to evaluate whether umpires made correct calls in terms of calling each pitch a strike or a ball.

To examine cold hand bias, we test whether the umpire is more likely to call the current pitch a strike if the most recent previous pitch was called a strike, controlling for the actual location of the pitch relative to the strike zone. We also test whether cold hand decisions are more likely to occur if there have been a streak in recent previous calls.

4.3 Indian Loan Officers

We use field experiment data collected by Cole et al. (forthcoming). The original intent of the experiment was to explore how various incentive schemes affect the quality of loan officer's assessment of loan quality. In the experiment, real loan officers are paid to screen actual loan applications under one of three randomly assigned incentive schemes: flat incentives that reward loan origination, stronger incentives which reward origination conditional on loan performance, and strongest incentives that reward origination of loans that default and penalize approval of loans that later default. In each session, each loan officer screens six randomly ordered loan files and decides whether to approve or reject each loan file.

The data also contains information on the underlying quality of the loan file (a continuous measure), loan officer background characteristics, a continuous score assigned by the loan officer reflecting his judgment of the quality of the file (this score does not affect experiment payout), as well as the time spent by the loan officer evaluating each loan file.

5 Baseline Estimation: Single Lag

We begin by focusing on binary decisions (*positive* = 1, *negative* = 0) on a sequence of cases. Is the current decision negatively correlated with the lagged decision?

$$Y_{it} = \beta_0 + \beta_1 Y_{i,t-1} + Controls + \epsilon_{it}$$

Cold hand bias predicts that $\beta_1 < 0$. Specifics pertaining to each of our three experiments are described below. In general, we include controls because:

1. There is likely to be heterogeneity across decision makers in their probability of making positive decisions. The tendency of a decision-maker to be positive could be a fixed characteristic or slowly changing over time. If we don't control for this, then both $Y_{i,t}$ and $Y_{i,t-1}$ will be correlated with this unobserved decision-maker characteristic, leading to an upward bias for β_1.

2. Decisions are also affected by underlying case quality. Ideally, the case quality of the current case is uncorrelated with $Y_{i,t-1}$. If there are moving average trends in case quality that we don't control for, β_1 will be upward biased. If case quality naturally has negative autocorrelation, β_1 will be downward biased.

5.1 Asylum Judges

- Y_{it} is an indicator for whether asylum is granted.
- Observations are at the judge x case order level. Cases are ordered within day and across days. Our sample includes observations in which the lagged case was viewed in the same day or the previous workday (e.g. we include the observation if the current case is viewed on Monday

and the lagged case was viewed on Friday). Observations in which there is a longer time gap between the current case and the lagged case are excluded from the sample. Multiple decisions on a single litigant are treated as one decision as they tend to be all "grants" or all "denies". Multiple family members are also treated as 1 observation for the same reason. We infer shared family status if cases share date, nationality, city, decision, presence of representation, and case type.

- *Controls* include:
  - A set of dummies of for the number of yes decisions over the past 6 decisions (excluding the current decision) of the judge. This controls for recent trends in grants, case quality, or judge mood.
  - A set of dummies of for the number of yes decisions over the past 6 decisions (excluding the current decision) across all judges in the city. This controls for recent trends in grants, case quality, or city mood.
  - We don't include judge FE because that automatically induces negative correlation between Y_{it} and $Y_{i,t-1}$.
  - Characteristics of current case: Presence of lawyer representation dummy, torture dummy, defensive case dummy, family size, etc.

5.2 Loan Officers

- Y_{it} is an indicator for whether the loan is approved.
- Observations are at the loan officer x loan order level. Loans are ordered within session and across sessions. Our sample includes observations in which the lagged loan was viewed in the same session (so we exclude the first loan viewed in each session because we do not expect cold hand bias across sessions which are often separated by multiple days).
- *Controls* include:
  - 5-part spline in the mean loan officer approval rate within each incentive treatment (calculated excluding the six observations corresponding to the current session): This

flexibly controls for loan officer x incentive scheme level approval rate. We don't include loan officer FE is because that automatically induces negative correlation between Y_{it} and $Y_{i,t-1}$. We don't use any observations in the current session to calculate the mean because loan quality within each session is balanced.

  - Mean approval rate of the current loan file (calculated excluding the current observation): This controls for the quality of the current loan and addresses the issue that our sample size is small and loan quality within each session is balanced.

- We will also split the sample by incentive type

  1. flat incentives
  2. reward if good loans are approved
  3. reward if good loans are approved and sharp punishment if bad loans are approved

5.3 Baseball Umpires

- Y_{it} is an indicator for whether the current pitch is called a strike
- $Y_{i,t-1}$ is an indicator for whether the previous pitch that was called was called a strike
- The sample includes all called pitches except for the first in each game
- *Controls* include:
  - count
  - home team dummy
  - pitch location dummies (a dummy for each 3x3 inch square grip inside and outside the strike zone)
  - linear controls for velocity, vertical movement, horizontal movement, and other Pitch F/X variables
- In this experiment, we are particularly concerned that the "quality" of the pitch will also display cold hand or react to the umpire's previous call. We estimate a version of the analysis

where Y (current and lagged) refers to scaled distance from the center of the strike zone. This tests for cold hand in pitch quality. We estimate another version of the analysis in which distance from the center of the strike zone is regressed on whether the lagged pitch was called a strike.

6 Streaks

How is the current decision related to streaks in past decisions? For now, we restrict our analysis to the last two decisions.

$$Y_{it} = \beta_0 + \beta_1 I(1,1) + \beta_2 I(1,0) + \beta_3 I(0,0) + Controls + \epsilon_{it}$$

Here, $I(1,1)$ is an indicator for the last two decisions being positive. $I(1,0)$ is an indicator for $Y_{i,t-2} = 1$ and $Y_{i,t-1} = 0$. $I(0,0)$ is an indicator for the last two decisions being negative. The omitted group is $I(0,1)$, an indicator for $Y_{i,t-2} = 0$ and $Y_{i,t-1} = 1$. If there is cold hand bias, we expect $\beta_1 < 0$, $\beta_2 > 0$, $\beta_3 > 0$, and $\beta_3 > \beta_2$. All controls are as described in the baseline specification. However, we restrict our sample so that the current decisions and as well as the two most recent decisions are consecutive (e.g. not broken across games, gaps in calendar days, or sessions).

7 Magnitudes

We estimate the fraction of decisions that would have been reversed if the decision maker had not suffered from cold hand bias. 1.5% of refugee decisions would have been decided the other way. By way of comparison, the average grant rate is 39% and the standard deviation of judges' mean grant rate is 19 percentage points. In other preliminary analysis of baseball umpires and loan officers, we find that up to 5% of decisions would have been reversed.

8 Heterogeneity and Interpretation Issues

We're interested in what types of decision makers or situations correspond to stronger cold hand bias. We would also like to distinguish between whether cold hand bias is caused by the internal desire to be fair or by the desire to appear fair to others. Within each experiment, we explore heterogeneity by:

1. Asylum judges: experience, higher caseload
2. Loan officers: incentive treatment, age, education, experience, time spent reviewing case
3. Baseball umpires: whether the game is important, umpire experience, home vs. away game, whether the inning is important in terms of determining the game outcome

9 Measurement Issues

9.1 Distinguishing cold hand bias from sequential contrast effects

Sequential contrast effects describes situations in which the decision maker's criteria for quality while judging the current case is higher if the previous case was particularly high quality. For example, consider a decision maker who judges physical beauty with a binary "hot" or "not" rating. After seeing a particularly attractive person (and deciding "hot"), the decision maker may be more likely to judge the next case "not" because her standard for beauty has been raised. Like cold hand bias, sequential contrast effects can lead to negative autocorrelation in decisions.

We distinguish cold hand bias from sequential contrast effects. First, sequential contrast effects are unlikely to occur in the context of baseball umpires.

Second, we can estimate:

$$Y_{it} = \beta_0 + \beta_1 Y_{i,t-1} + \beta_2 Quality_{i,t-1} + Controls + \epsilon_{it}$$

If sequential contrast drives the cold hand bias, then we expect to find that $\beta_2 < 0$. Controlling for the discrete decision $Y_{i,t-1}$, judges should be more likely to reject the current case if the previous case was of high quality, as measured continuously using $Quality_{i,t-1}$. This test assumes that our measure of lagged quality is a comprehensive measure of true quality.

10 Results

10.1 Asylum Judges

In Table 1, column 1, we show that an asylum denial is 1.5% more likely if the previous grant was an approve rather than deny. We control for applicant characteristics and a measure of a judge's time-varying approval rate, a set of dummies for the number of grants out of the previous 5 decisions (not including the current decision), and a measure of a city's time-varying approval rate, a set of dummies for the number of grants out of the previous 5 decisions across all judges in the city. In column 2, we also control for judge-specific time trends. Column 3 controls for time-of-day fixed effects.

Table 1
Baseline

Dependent Variable	Grant		
	(1)	(2)	(3)
Lagged Grant	-0.0159***	-0.0116***	-0.0156***
	(0.00422)	(0.00401)	(0.00422)
Applicant Controls	Yes	Yes	Yes
Num prev asylums granted by judge	Yes	Yes	Yes
Num prev asylums granted in city	Yes	Yes	Yes
Judge-specific time trends	No	Yes	No
Time of day	No	No	Yes
N	106071	106071	106071
R^2	0.125	0.167	0.126

Notes: Standard errors in parentheses (* p < 0.10; ** p < 0.05; *** p < 0.01). This table tests whether the decision to grant asylum for the current applicant is related to the decision to grant asylum to the previous applicant. Observations are at the judge x applicant level. Observations are restricted to decisions that occurred within one day or weekend after the previous decision. Number of previous asylums granted is the full set of fixed effects for number of grants out of the previous 5 decisions to control for a time-varying grant tendency of a judge in the recent time period. Number of previous asylums granted in the city is the same at the city level. Judge-specific time trends are a linear time trend specific to a judge. Time of day are fixed effects for the start time of the hearing. Standard errors are clustered by judge.

Table 2, column 1, shows that caseload exacerbates cold hand bias. If there is only 1 case per day, there is no cold hand bias with respect to the previous decision. However, each additional case per day corresponds to an additional 2.6% in cold hand bias. The average number of asylum cases that a judge handles per day is 1.2. The next two columns address the concern that we do not have measures of applicant quality, which may be negatively serially correlated at the city level. We have cases where the judge tele- or video-conferences, where the hearing is in a different location from the courthouse (for example, the hearing may be in Anchorage when the judge is in Portland or Seattle). We find that cold hand bias is 4.9% greater for these decisions (column 2). This suggests that any correlation in consecutive case quality within the same city is actually likely to be positive, and therefore a bias against our findings. Column 3 shows that cold hand bias is also greater, though not significantly so, when the judge's previous decision is in a different city. Next we show that moderate judges (those who make 30-70% decisions as grants on other days) display more cold hand bias (column 4). Finally, we show that the cold hand bias is 4% greater when the previous decision is on the same day.

Table 2
Heterogeneity

Dependent Variable	Grant				
	(1)	(2)	(3)	(4)	(5)
Lagged Grant	0.0239***	-0.0137***	-0.0152***	0.0109	-0.00184
	(0.00825)	(0.00428)	(0.00435)	(0.0112)	(0.00440)
Caseload	0.0179***				
	(0.00361)				
Lagged Grant * Caseload	-0.0259***				
	(0.00499)				
Hearing not at base city		0.00447			
		(0.0179)			
Lagged Grant * Hearing not at base city		-0.0491***			
		(0.0126)			
Previous decision in different city			-0.00720		
			(0.00930)		
Lagged Grant * Previous decision in different city			-0.0221		
			(0.0235)		
Moderate Judge				0.129***	
				(0.00923)	
Lagged Grant * Moderate Judge				-0.0428**	
				(0.0193)	
Previous decision on same day					0.0340***
					(0.00433)
Lagged Grant * Previous decision on same day					-0.0403***
					(0.00721)
Applicant controls	Yes	Yes	Yes	Yes	Yes
Num prev asylums granted by judge	Yes	Yes	Yes	Yes	Yes
Num prev asylums granted in city	Yes	Yes	Yes	Yes	Yes
N	106071	106071	106071	106071	106071
R^2	0.125	0.125	0.125	0.138	0.126

Notes: Standard errors in parentheses (* p < 0.10; ** p < 0.05; *** p < 0.01). Column 1 tests whether cold hand bias is weaker when judges have a lower case load. "Caseload" is the number of cases on asylum, asylum-witholding, or withholding-convention against torture on a day that a judge makes a decision on. Column 2 tests whether cold hand bias is stronger when possible omitted serial correlation in applicant quality within a city is absent, i.e. when a judge hears a case in a different city (by teleconference or videoconference). Column 3 does the same for the previous decision being in a different city. Column 4 tests whether cold hand bias is larger for moderate judges (those whose average grant rate outside today is between 30% and 70%). Column 5 tests whether cold hand bias is worse when the previous decision is on the same day. All other variables and restrictions are as described in Table 1. Standard errors are clustered by judge.

Next, we show that after a streak of two denials, judges are 1.8% more likely to grant. Following a grant then deny decision, the judge is also 1.7% more likely to grant relative to a judge who denied and then granted (Table 3). The sample is smaller because both of the previous two decisions by the judge has to satisfy the condition that the prior decision is within one day of the current decision, or within one weekend if the current decision is a Monday decision.

Table 3
Reaction to Streaks by Asylum Judges

Dependant Variable	Grant
	(1)
Grant-Grant	-0.00770
	(0.00946)
Deny-Grant	0.0174***
	(0.00605)
Deny-Deny	0.0176***
	(0.00663)
Applicant controls	Yes
Num prev asylums granted by judge	Yes
Num prev asylums granted in city	Yes
N	46497
R^2	0.121

Notes: Standard errors in parentheses (* p < 0.10; ** p < 0.05; *** p < 0.01). This table tests how judges react to streaks in past decisions. Grant-Grant is a dummy equal to 1 if the judge approved the two most recent asylum applicants. Deny-Grant is a dummy equal to 1 if the judge denied the most recent applicant and granted the applicant before that. Deny-Deny is a dummy equal to 1 if the judge denied the last two applications. The sample restricts to decisions where the current and previous decision both satisfy the requirement of occurring within one day or weekend after its previous decision. All other variables and restrictions are as described in Table 1. Standard errors are clustered by judge.

Finally, Table 4 shows that judges who are inexperienced (less than the median experience of 8 years) are particularly likely to display cold hand bias. Cold hand bias of experienced judges is statistically indistinguishable from 0. The sample is slightly smaller because we do not have the entire set of biographies.

10.2 Loan Officers

Table 5 shows that for the entire sample, if the previous decision was an approve, then the next decision is 6% points more likely to be deny. This effect is strongest among loan officers with flat incentives (14% points) and weaker with incentives.

Table 4
Asylum Judge Experience

Dependant Variable	Grant
	(1)
Lagged Grant	-0.0246***
	(0.00806)
Experienced	0.00250
	(0.00827)
Lagged Grant * Experienced	0.0291**
	(0.0127)
Applicant controls	Yes
Number of previous asylums granted judge	Yes
Number of previous asylums granted city	Yes
N	80336
R^2	0.087

Notes: Standard errors in parentheses (* $p < 0.10$; ** $p < 0.05$; *** $p < 0.01$). This table tests how cold hand bias differs by judge experience. Experience is defined as above median experience of 8 years. All other variables and restrictions are as described in Table 1. Standard errors are clustered by judge.

Table 5
Baseline

This table tests whether the decision to approve the current loan file is related to the decision to approve the previous loan file. Observations are at the loan officer x loan file level and exclude (as a dependent variable) the first loan file evaluated within each experimental session. Column 1 includes the full sample while Columns 2-4 are restricted to sessions in which loan officers faced flat (20 Rs for each approved loan), stronger (20 Rs for each approved loan that does not default and 10 Rs for each declined loan), and strongest incentives (20 Rs for each approved loan that does not default and -100 Rs for each approved loan that does default), respectively. Other control variables include "Loan Officer Trend," the loan officer's average approval rate for loans in other sessions within the same incentive treatment, excluding the current session (introduced as a 5-part spline) and "Loan Quality," two dummy variables for whether, in a real world setting, the loan file was rejected or approved but defaulted. Standard errors are clustered by loan officer. *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

	All	Flat Incentives	Stronger Incentives	Strongest Incentives
	(1)	(2)	(3)	(4)
Lagged Approve	-0.0612***	-0.146***	-0.0599***	-0.0266
	(0.0130)	(0.0343)	(0.0145)	(0.0326)
Loan Officer Trend	Yes	Yes	Yes	Yes
Loan Quality	Yes	Yes	Yes	Yes
N	6650	850	4675	1125
R^2	0.0650	0.0768	0.0684	0.0612

Table 6 shows that moderate loan officers have roughly twice the level of cold hand bias. The less time spent viewing the loan file, the more cold hand bias (Panel B).

Table 6

Heterogeneity

Panel A tests whether cold hand bias is stronger when the loan officer has a moderate rather than extreme overall loan approval rate. "Moderate" in a dummy equal to 1 if the loan officer's average approval rate for loans, excluding the current session, is between 0.30 and 0.70 inclusive. Panel B tests whether cold hand bias is weaker when loan officers spend more time reviewing the current loan file. "Time Viewed" is the number of minutes spent reviewing the current loan file. All other variables are as described in Table 1. Standard errors are clustered by loan officer. *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

Panel A: Moderate vs. Extreme Loan Officers

	All	Flat Incentives	Stronger Incentives	Strongest Incentives
	(1)	(2)	(3)	(4)
Lagged Approve	-0.0426***	-0.0967***	-0.0486***	-0.00383
	(0.0154)	(0.0349)	(0.0177)	(0.0419)
Moderate	-0.00845	0.0232	-0.0108	-0.0106
	(0.0248)	(0.0542)	(0.0284)	(0.0507)
Lagged Approve x Moderate	-0.0505*	-0.175**	-0.0300	-0.0615
	(0.0265)	(0.0665)	(0.0297)	(0.0616)
Loan Officer Trend	Yes	Yes	Yes	Yes
Loan Quality	Yes	Yes	Yes	Yes
N	6650	850	4675	1125
R^2	0.0669	0.0922	0.0693	0.0639

Panel B: Time Spent Reviewing Loan File

	All	Flat Incentives	Stronger Incentives	Strongest Incentives
	(1)	(2)	(3)	(4)
Lagged Approve	-0.115***	-0.166***	-0.107***	-0.154***
	(0.0204)	(0.0425)	(0.0233)	(0.0542)
Time Viewed	-0.0166***	-0.0182	-0.0152***	-0.0290**
	(0.00389)	(0.0125)	(0.00453)	(0.0126)
Lagged Approve x Time Viewed	0.0161***	0.00673	0.0135**	0.0441***
	(0.00439)	(0.0124)	(0.00524)	(0.0138)
Loan Officer Trend	Yes	Yes	Yes	Yes
Loan Quality	Yes	Yes	Yes	Yes
N	6650	850	4675	1125
R^2	0.0673	0.0808	0.0707	0.0690

When loan officers approve two applications in a row, the next decision is 14% more likely to be a deny, relative to when the loan officer denied two applications in a row (Table 7). After an approval, then rejection, the next decision is 5% more likely to be a rejection relative to when the officer made two rejections in a row.

Table 7

Reaction to Streaks

This table tests how loan officers react to streaks in past decisions. Approve-Approve is a dummy equal to 1 if the loan officer approved the two most recent previous loans. Approve-Reject is a dummy equal to 1 if the loan officer approved the most recent pervious loan and rejected the loan before that. Reject-Approve is a dummy equal to 1 if the loan officer rejected the most recent previous loan and approved the loan before that. The omitted category is Reject-Reject, which is a dummy equal to 1 if the loan officer rejected the two most recent previous loans. The sample excludes observations corresponding to the first two loans reviewed within each session. All other variables are as described in Table 1. Standard errors are clustered by loan officer. *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

	All	Flat Incentives	Stronger Incentives	Strongest Incentives
	(1)	(2)	(3)	(4)
Approve-Approve	-0.143***	-0.126	-0.181***	-0.0527
	(0.0247)	(0.0871)	(0.0283)	(0.0497)
Approve-Reject	-0.0139	-0.0194	-0.0374	0.0384
	(0.0246)	(0.0961)	(0.0265)	(0.0515)
Reject-Approve	-0.0504**	0.0390	-0.0900***	0.00119
	(0.0253)	(0.0917)	(0.0274)	(0.0591)
Loan Officer Trend	Yes	Yes	Yes	Yes
Loan Quality	Yes	Yes	Yes	Yes
N	5320	680	3740	900
R^2	0.0760	0.0848	0.0867	0.0607

Table 8 shows the effect is robust to controlling for the number of previous loans approved.

Table 8

Recency Effect

This table tests whether loan officers react to the most recent decision after controlling for the total number of previously approved loans within the same session. "Num Previous Loans Approved" represents a set of dummy variables for the number of previously approved loans. For example, suppose the loan officer is currently reviewing loan file number 4 and has previously approved one loan of the three reviewed so far within the session. We test whether the loan officer is less likely to approve the current loan if she approved the most recent loan file rather than if she had approved an earlier loan among the three loans reviewed so far. The sample excludes observations corresponding to the first two loans reviewed within each session as the dependent variable. All other variables are as described in Table 1. Standard errors are clustered by loan officer. *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

	All	Flat Incentives	Stronger Incentives	Strongest Incentives
	(1)	(2)	(3)	(4)
Lagged Approve	-0.0248	-0.0862**	-0.0249	-0.00120
	(0.0151)	(0.0358)	(0.0166)	(0.0424)
Loan Officer Trend	Yes	Yes	Yes	Yes
Loan Quality	Yes	Yes	Yes	Yes
Num previous loans approved	Yes		Yes	Yes
N	5320	680	3740	900
R^2	0.0911	0.148	0.0989	0.0724

Table 9 shows that lagged approval still predicts a 5.5% increase in rejection even after controlling for lagged loan quality. This suggests that sequential contrast effect is not driving the result, as the lagged loan quality should have a strong negative effect on the next decision. In fact, when self-reported scores are used, the cold hand bias coefficient becomes even larger than when lagged loan quality is not included.

Table 9

Cold Hand vs. Sequential Contrast Effects

This table tests whether the negative correlation between current loan approval and lagged loan approval could be caused by sequential contrast effects. "Lagged Loan Quality" is a continuous measure of the quality of the most recently reviewed loan file while Lagged Approve is a binary measure of whether the previous loan was approved. Conditional on the binary measure of whether the previous loan was approved, sequential contrast effects predict that the loan officer should be less likely to approve the current loan if the previous loan was of higher quality, measured continuously. In other words, sequential contrast effects predicts that the coefficient on "Lagged Loan Quality" should be negative. In Panel A, loan quality is measured as the standardized average approval rate of the loan file by other loan officers, excluding the loan officer corresponding to the current observation. In Panel B, loan quality is measured as the loan officer's own self reported loan rating (on a scale from 20-100). Both measures of loan quality are standardized to have a mean of 0 and a standard deviation of 1. Standard errors are clustered by loan officer. *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

Panel A: Loan Quality = Loan File Approval Rate (Leave-Out-Mean)

	All	Flat Incentives	Stronger Incentives	Strongest Incentives
	(1)	(2)	(3)	(4)
Lagged Approve	-0.0546***	-0.143***	-0.0533***	-0.0183
	(0.0139)	(0.0369)	(0.0157)	(0.0335)
Lagged Loan Quality	-0.00906	-0.000540	-0.00932	-0.0141
	(0.00580)	(0.0163)	(0.00720)	(0.0144)
Loan Officer Trend	Yes	Yes	Yes	Yes
Loan Quality	Yes	Yes	Yes	Yes
N	6634	847	4665	1122
R^2	0.0655	0.0753	0.0691	0.0616

Panel B: Loan Quality = Self Reported Score

	All	Flat Incentives	Stronger Incentives	Strongest Incentives
	(1)	(2)	(3)	(4)
Lagged Approve	-0.0704***	-0.152***	-0.0684***	-0.0506
	(0.0138)	(0.0359)	(0.0152)	(0.0353)
Lagged Loan Quality	0.00913	0.00616	0.00824	0.0244
	(0.00584)	(0.0203)	(0.00666)	(0.0188)
Loan Officer Trend	Yes	Yes	Yes	Yes
Loan Quality	Yes	Yes	Yes	Yes
N	6650	850	4675	1125
R^2	0.0653	0.0769	0.0687	0.0627

Table 10 shows that cold hand bias is strongest in the flat incentive treatment.

Table 10

Detailed Incentive Breakdown

This table tests how cold hand bias differs across six incentive treatments. Each of the three incentive treatments in Table 1 is divided into a standard and optional information treatment. In the optional information treatment, loan officers receive an initial information endowment of Rs 108 (US$ 2.25) which they may spend to view additional sections of the loan file. All other variables are as described in Table 1. Standard errors are clustered by loan officer. *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

	Flat	Stronger	Strongest	Flat + Info	Stronger + Info	Strongest + Info
	(1)	(2)	(3)	(4)	(5)	(6)
Lagged Approve	-0.196***	-0.0509***	-0.0588	-0.114***	-0.0747***	-0.00638
	(0.0463)	(0.0169)	(0.0476)	(0.0415)	(0.0205)	(0.0407)
Loan Officer Tred	Yes	Yes	Yes	Yes	Yes	Yes
Loan Quality	Yes	Yes	Yes	Yes	Yes	Yes
N	445	2770	505	405	1905	620
R^2	0.0843	0.0663	0.0590	0.106	0.0731	0.0717

Table 11 and 12 show that judges with experience and graduate school education are less likely to have cold hand bias. The quantity reduction in cold hand bias is economically large but statistically insignificant given the small sample size.

Table 11
Loan Officer Experience

	All	Flat Incentives	Stronger Incentives	Strongest Incentives
	(1)	(2)	(3)	(4)
Lagged Approve	-0.0714***	-0.100**	-0.0798***	-0.0370
	(0.0187)	(0.0452)	(0.0186)	(0.0474)
Experienced	-0.0249	0.0681	-0.0333	-0.0456
	(0.0201)	(0.0487)	(0.0210)	(0.0582)
Lagged Approve x Experienced	0.0154	-0.0831	0.0285	0.0163
	(0.0249)	(0.0645)	(0.0261)	(0.0618)
Loan Officer Trend	Yes	Yes	Yes	Yes
Loan Quality	Yes	Yes	Yes	Yes
N	6650	850	4675	1125
R^2	0.0653	0.0785	0.0688	0.0624

Table 12
Loan Officer Education

	All	Flat Incentives	Stronger Incentives	Strongest Incentives
	(1)	(2)	(3)	(4)
Lagged Approve	-0.0747***	-0.158***	-0.0711***	-0.0571
	(0.0155)	(0.0370)	(0.0170)	(0.0360)
Grad School	-0.0143	0.00890	-0.0163	-0.0244
	(0.0204)	(0.0620)	(0.0250)	(0.0529)
Lagged Approve x Grad School	0.0474*	0.0380	0.0394	0.112
	(0.0253)	(0.0795)	(0.0297)	(0.0680)
Loan Officer Trend	Yes	Yes	Yes	Yes
Loan Quality	Yes	Yes	Yes	Yes
N	6650	850	4675	1125
R^2	0.0657	0.0784	0.0688	0.0652

10.3 Baseball Umpires

Table 13 Column 1 shows that umpires are 1 percentage point less likely to call a pitch a strike if the most recent previously called pitch was also called a strike. Column 2 shows that cold hand bias is stronger following streaks. Umpires are 1.3 percentage points more likely to call a pitch a strike if the two most recent called pitches were also called strikes. Further, umpires are more likely to call the current pitch a strike if the most recent pitch was called a strike and the pitch before that was called a ball than if the ordering of the last two calls were reversed. In other words, recency matters. All analysis in this and subsequent tables include detailed controls for the actual location, speed, and curvature of the pitch.

Table 13
Called Pitches – Baseball Umpires

1 Y=Prob.(strike) on X with location controls

1.1 Called Pitches - Strike - All Controls

Number of previous pitches	1	2
All Strikes	-0.00913^{***} (0.000598)	-0.0131^{***} (0.00107)
Ball-Strike	—	-0.00272^{***} (0.000652)
Strike-Ball	—	-0.00989^{***} (0.00072)
N	1522465	1319542
R^2	0.661	0.66
Adj R^2	0.661	0.659

Table 1: *Regression controlling for count, home team, 3 inch square grids inside and outside the strike zone, movement, other Pitch F/X variables, and the proportion of previous pitches faced by a batter in a given game that fall into one of seven categories with a left hand variable of prob.(strike)*

Table 14 repeats the above analysis but restricts the sample to pitches that were called consecutively (so both the current and most recent pitch received umpire calls). In this restricted sample, the umpire's recent previous calls may be more salient because they are not separated by uncalled pitches. We find that the magnitude of the cold hand bias increases significantly in this sample.

Table 14
Consecutive Pitches – Baseball Umpires

1.2 Consecutive Pitches - Strike - All Controls

Number of previous pitches	1	2
All Strikes	-0.0142^{***} (0.001)	-0.021^{***} (0.00288)
Ball-Strike	—	-0.00707^{***} (0.00155)
Strike-Ball	—	-0.0187^{***} (0.00156)
N	891251	424706
R^2	0.657	0.661
Adj R^2	0.656	0.66

Table 2: *Regression controlling for count, home team, 3 inch square grids inside and outside the strike zone, movement, other Pitch F/X variables, and the proportion of previous pitches faced by a batter in a given game that fall into one of seven categories with a left hand variable of prob.(strike)*

Tables 15 and 16 test that the negative autocorrelation in umpire calls is not due to bias caused by changes in the actual location of the pitch. We repeat the previous analysis but use distance from the center of the strike zone as our dependent variable. If pitchers are more likely to throw true balls after the previous pitch was called a strike, we should find significant negative coefficients.

Instead we find small coefficients that are insignificantly different from zero.

Table 15
Called Pitches - Strikes – Baseball Umpires

3 Y=distance on X with location controls

3.1 Called Pitches - Strike - All Controls

Number of previous pitches	1	2
All Strikes	−0.000466 (0.000818)	−0.00109 (0.00132)
Ball-Strike	—	−0.00138 (0.000805)
Strike-Ball	—	−0.000742 (0.000888)
N	1522465	1319542
R^2	0.724	0.762
Adj R^2	0.724	0.761

Table 5: *Regression controlling for count, home team, 3 inch square grids inside and outside the strike zone, movement, other Pitch F/X variables, and the proportion of previous pitches faced by a batter in a given game that fall into one of seven categories with a left hand variable of the scaled score for the distance from the center of the strike zone*

Table 16
Consecutive Pitches - Strikes – Baseball Umpires

3.2 Consecutive Pitches - Strike - All Controls

Number of previous pitches	1	2
All Strikes	0.0000607 (0.00149)	−0.00436 (0.00388)
Ball-Strike	—	0.00141 (0.00208)
Strike-Ball	—	0.00137 (0.00209)
N	891251	424706
R^2	0.667	0.705
Adj R^2	0.666	0.704

Table 6: *Regression controlling for count, home team, 3 inch square grids inside and outside the strike zone, movement, other Pitch F/X variables, and the proportion of previous pitches faced by a batter in a given game that fall into one of seven categories with a left hand variable of the scaled score for the distance from the center of the strike zone*

In preliminary tests (unreported) we also test whether our results could be driven by sequential contrast effects (SCE). SCE would predict that the current pitch would be more likely to be called a strike if the previous pitch was very "high quality," i.e. obviously not a strike. We find the opposite. Negative autocorrelation is stronger when the previous called pitch was near the edge of the strike zone. In other words, negative autocorrelation is stronger when the previous case was of ambiguous quality.

11 Conclusion

Decision-makers operating under uncertainty about what is fair and just may worry that she is becoming or appearing too lenient or tough if she issues a streak of yes or no decisions. Previous research on the hot hand fallacy suggests that agents often think of sequential streaks of 0's or 1's as evidence of bias even though such streaks are likely to occur through flips of a fair coin. We hypothesize that the hot hand fallacy leads agents to engage in cold hand, i.e. negatively autocorrelated, decision-making. If cases are ordered randomly, a judge's decision on the previous case should not predict her decision on the next case if decisions are made based upon case merits. However, a judge may be worried that she is becoming or appearing too lenient or tough if she issues a streak of affirmative or negative decisions, and may therefore actively adjust her decisions in the opposite direction. We test the cold-hand hypothesis in three contexts: judicial decisions in U.S. refugee asylum cases, umpire calls on baseball pitches, and loan officer loan approval decisions. We find evidence of negative autocorrelation, particularly among moderate decision-makers who may be ambiguity averse, and stronger effects after recent streaks of decisions and in situations when agents face weak incentives for accuracy. We also show that our findings are unlikely to be driven by sequential contrast effects. In ongoing work, we explore the welfare consequences of these judicial biases by linking the refugee court decisions to their possible appeals.

References

Andreoni, James and B. Douglas Bernheim, "Social Image and the 50-50 Norm: A Theoretical and Experimental Analysis of Audience Effects," *Econometrica*, September 2009, *77* (5), 1607–1636.

Chen, Daniel L. and Carlos Berdejó, "Priming Ideology? Electoral Cycles without Electoral Incentives among Elite U.S. Judges," Technical Report, ETH Zurich, Mimeo 2013.

_ **and Martin Schonger**, "Social Preferences or Sacred Values? Theory and Evidence of Deontological Motivations," Working Paper, ETH Zurich, Mimeo August 2013.

Cole, Shawn, Martin Kanz, and Leora Klapper, "Incentivizing Calculated Risk-Taking: Evidence from an Experiment with Commercial Bank Loan Officeres," *Journal of Finance*, forthcoming.

Fehr, Ernst, Lorenz Goette, and Christian Zehnder, "A Behavioral Account of the Labor Market: The Role of Fairness Concerns," *Annual Review of Economics*, 2009, *1* (1), 355–384.

Gilovich, Thomas, Robert Vallone, and Amos Tversky, "The Hot Hand in Basketball: On the Misperception of Random Sequences," *Cognitive Psychology*, 1985, *17*, 295–314.

Guthrie, Chris, Jeffrey J. Rachlinski, and Andrew J. Wistrich, "Inside the Judidical Mind," *Cornell Law Review*, 2000, *86* (4), 777–830.

Krosnick, Jon A. and Donald R. Kinder, "Altering the Foundations of Support for the President Through Priming," *The American Political Science Review*, 1990, *84* (2), 497–512.

Levitt, Steven and John List, "What Do Laboratory Experiments Tell Us About the Real World?," *The Journal of Economic Perspectives*, 2007, *21* (2), 153–174.

Malkiel, Burton G., "Returns from Investing in Equity Mutual Funds 1971 to 1991," *The Journal of Finance*, 1995, *50* (2), 549–572.

Mas, Alexandre, "Pay, Reference Points, and Police Performance," *The Quarterly Journal of Economics*, 2006, *121* (3), 783–821.

Rabin, Matthew, "Incorporating Fairness into Game Theory and Economics," *The American Economic Review*, December 1993, *83* (5), 1281–1302.

_ , "Inference by Believers in the Law of Small Numbers," *The Quarterly Journal of Economics*, August 2002, *117* (3), 775–816.

Tversky, Amos and Daniel Kahneman, "Judgment under Uncertainty: Heuristics and Biases," *Science*, 1974, *185* (4157), 1124–1131.