A Online Appendix

In § A.1, we discuss a rational setting with partial communication of past signals. §A.2 provides discussion related to Section 6 on limited communication or observation of past actions. §A.3 discusses legal precedent and information cascades, relating to the discussion in §10.6. §A.4 describes a simple model of belief cascades, relating to the discussion in §10.8. §A.5 discusses how observation of others can degrade decision quality, relating to the discussion in §7.2.

A.1 Recent past signals observable

The SBM and the BHW model are based on the premise that agents do not directly communicate their private signals. If the signals of agents are fully communicated to their followers, learning would be efficient, and there would be asymptotic learning.

In many practical contexts, more information is indeed transmitted than just action choices. For example, people are often free to talk about their private information, though owing to time and cognitive constraints, such communication may be imperfect. This raises the question of whether incorrect cascades still occur, and whether there is asymptotic learning.

We consider here the possibility of limited communication of private signals. Every private signal is directly communicated, which in principle could bring about asymptotic learning. However, we consider a case in which each signal is passed on to just one other agent, and provide an example in which incorrect cascades still form and last forever.

Consider the SBM of Section 2, except that in addition to observing all past actions, each agent observes the private signal of the agent's immediate predecessor. In contrast with the SBM, we assume that when indifferent, each agent always follows her own signal. This assumption is convenient but not crucial.

After actions HH or LL, I_3 infers that the first two signals were hh or $\ell\ell$ respectively, and falls into a cascade which could easily be incorrect. This cascade can be broken. To see this, suppose that the first two actions are HH, and that I_3 observes ℓ. So I_3 chooses H. If I_4 observes ℓ, this combines with observation of I_3's ℓ to make I_4 indifferent, so I_4 rejects, breaking the cascade.

Consider next the case in which I_3 observes h as well. Now, I_4 adopts, because I_4 knows that the first 3 signals were all h. At this point, the cascade of adoption continues forever. I_5 knows that the first two signals were hh, and that the third and fourth signals were not $\ell\ell$. (I.e., they were either ℓh, hh, or $h\ell$.) So the net evidence in favor of state H is at least slightly stronger than two h signals. So even in the worst case in which I_5 directly observes two ℓ signals (I_5's and I_4's signals), I_5 strictly prefers H. Similar reasoning shows that all subsequent agents adopt.

The general point that inefficient cascades can form and last forever does not require this tie-breaking convention.[1] Intuitively, the ability to observe the predecessor's signal is a double-edged blade. The immediate effect is to increase an agent's information, which increases the probability that an agent decides correctly. But to the extent that this is the case, it later makes the action history more informative. That eventually encourages later agents to fall into a cascade, which blocks asymptotic learning.

In particular, the action sequence becomes informative enough that agents fall into cascades despite their access to an extra signal from the predecessor. The action history eventually overwhelms an agent's information (even inclusive of observation of a predecessor's signal), at which point the accuracy of the social belief stops growing.

Another way to think of this is that being able to observe predecessors' signals is somewhat like being able to observe multiple private signals instead of one. This

[1] The following modification illustrates the point in a setting with no ties. Suppose that agents differ very slightly in their precisions, and that agents act in inverse order of their precisions, from low to high. Now agents are never indifferent. Agents who are close to indifference strictly prefer to follow their own signals, and the same analysis holds.

does not fundamentally change the argument for why incorrect cascades form—that the accumulation of information implicit in past actions must eventually overwhelm the signal(s) of a single agent.

A.2 Further discussion of §6 on limited communication or observation of past actions

In the next two subsections, we discuss imperfectly rational models, as referred to in § 6.2.3, that make direct heuristic assumptions about the agent's mapping from observed actions and payoffs into the agent's actions.

A.2.1 Heuristic action rules based on observation of averages of past payoffs and of action frequency

If rational agents could observe the average payoffs experienced from possible actions undertaken by many in a large population, they would always choose the right action. However, it is possible that agents fall short of rationality. In the model of Ellison and Fudenberg (1993), there is a continuum of agents, where each period some given fraction of them has the opportunity to update their choices. Agents observe the average payoff of the two actions in the population from the previous period, and their popularities. Agents use rules of thumb for updating their actions, which take the form of a specified probability of switching actions based on these two variables. The paper explores conditions on the weighting of these considerations that promote or hinder long-run correct decisions.

A.2.2 Heuristic action rules based on sampling of past actions and payoffs

In Ellison and Fudenberg (1995), there is a large number of agents who take simultaneous actions at each discrete period. Each period, some fraction of agents exogenously stick to their current action, and the remainder observes an unbiased sample of the latest actions and payoffs of N and choose actions based on this and based on their own latest payoff. Agents follow the following heuristic decision rule. If all the agents in an agent's sample take the same action, the agent follows that action. If both actions are selected by at least one agent in the sample, the agent chooses whichever action has a higher average payoff based on observed reports and the agent's own latest experienced payoff.

When the sample size N is small, learning from the experience of others causes the system to evolve to universal adoption of the correct action. In contrast, when the sample size is large, there is strong pressure toward diversity of behavior.[2] So the system never fixes on the correct action. This analysis is valuable in illustrating how non-obvious interesting conclusions about efficiency derive from reasonably plausible heuristic assumptions. On the other hand, when there is a 'split decision' in an agent's sample, it would be reasonable for an agent to take into account that a preponderance of 99 Adopts to 1 Reject (for example) might suggest almost conclusively that adopt is superior.[3] This would tend to oppose the diversity effect discussed above.

A.3 Legal Precedent and Information Cascades

§ 10.6 discusses the application of cascades to legal precedent. In the models of Talley (1999) and Daughety and Reinganum (1999), this takes the form of imitation of decisions across courts (see also Vermeule (2012)). Daughety and Reinganum (1999) model imitation in which courts at the same level observe private signals about a binary state. This state indicates the single correct decision for a set of related cases to be considered by these courts. (The context is appellate courts that receive related cases.) The courts act in exogenous sequence. If signals are bounded, imitation across the courts results

[2] This is because if one action is very unpopular, with large N many adherents of the more popular action "hear about" the unpopular one and potentially convert to it.

[3] Ellison and Fudenberg consider a related effect which they call 'popularity weighting.'

in information cascades. Even if signals are unbounded, there is a high probability of extensive imitation, resulting in poor information aggregation. The authors offer possible examples of actual precedential cascades among appellate courts.[4]

Talley argues that the conditions for cascades to occur in the precedential context are highly restrictive, in part because judges can relay granular information via written opinions. However, owing to limited time and attention, later courts may place heavier weight upon early decisions than upon the nuances of written opinions.

Daughety and Reinganum emphasize that incorrect cascades can be persistent, despite the finding of BHW that cascades are fragile, because of the rarity of shocks that might dislodge a judicial cascade. As discussed in § 2, informative public disclosures can dislodge cascades. However, the authors argue that in the U.S. judicial setting, such shocks take the form of cases being brought to the Supreme Court despite harmonious decisions of lower courts. Such review is rare, because the Supreme Court needs to wait for cases to be appealed, and because decisions that are harmonious across courts are rarely reviewed.

Sunstein (2009) argues that the adoption of legislation sometimes takes the form of information cascades across nations. Sunstein also argues that legal resolution of constitutional questions, such as the once-common, now rejected, view that the U.S. Constitution permits racial segregation, may be the product of information cascades.

[4]To the extent that *stare decisis* is mandated rather than just a description of court behavior, a court may feel pressured to follow precedent rather than doing so for purely informational reasons. However, courts do also deviate from precedent, perhaps because there can be gray areas in the applicability of a precedent. The pressure of *stare decisis* as a social norm presents a challenge for empirical testing for informational effects. At a deeper level, however, social learning may explain why the norm of *stare decisis* originally emerged in common law legal systems.

A.4 A Model of Belief Cascades

We provide here a simple model of belief cascades (cascades of reported beliefs), as described informally in §10.8.

In the SBM, agent I_n's belief $\hat{p}_n$ about the state is a continuous variable. Suppose that we extend the SBM by having each agent directly and truthfully communicate a report of her belief (probability assessment) to the next agent. Then each agent's private signal always influences the next agent's belief, so no agent is ever in a cascade of reported beliefs. However, in alternative settings, there can be cascades—including incorrect cascades—in reported beliefs. We focus on one possible source of belief cascades, based upon discreteness in the communication channel.

People often communicate in coarse categories, such as binary partitions.[5] Similarly, when asked for a suggestion about a movie or restaurant, or when discussing political topics, people often just name the preferred option, or say that an option is "hot" versus "sucks;" or "cool" versus "bogus."

To capture such limited bandwidth in the communication of beliefs, we consider a setting in which agents make decisions in a commonly known deterministic sequence, and each agent I_m observes the actions of some subset of the agent's predecessors.[6] Suppose that the action of each agent I_n is to report a binary indicator of state, $a_n \in \{L, H\}$.

We interpret the report a_n as the *reported belief* about the state. We will view reported belief L as a claim that the state is probably L, and reported belief H that the state is probably H.[7] This could take the form of expressing that "Candidate H will

[5] As the well-known hedge fund manager and self-improvement author Ray Dalio put it, "It is common for conversations to consist of people sharing their conclusions rather than exploring the reasoning that led to those conclusions."

[6] Alternatively, each agent might also recall and pass on the history of all beliefs that have been reported to that agent. If signals are discrete, such a setting would effectively be equivalent to the model of BHW.

[7] A possible behavioral extension of the model would be to have the reports sometimes understood

win the election" or “Candidate H will lose the election." Specifically, I_n *reports* belief L if her true belief $\hat{p}_n < 0.5$, reports H if $\hat{p}_n > 0.5$, and follows some indifference rule, such as flipping a coin between reporting L or H, if $\hat{p}_i = 0.5$. This reporting rule can be endogenized if agents desire a reputation with receivers for making good reports.

As defined in the main text, a *belief cascade* is a situation in which, having received the reported beliefs of some set of predecessors, an agent's reported belief is independent of the agent's private signal. A belief cascade is also an information cascade, as defined earlier, since the reported belief a_n can be interpreted as the agent's action.

In this model, if signals are bounded, then as in § 4, asymptotic learning can fail; even in the limit, reported and actual beliefs can be incorrect. Furthermore, if signals are also discrete, then for reasons similar to those of the model of BHW, there are incorrect belief cascades (see footnote 6).

In the modified SBM example above, belief cascades derive from communication bandwidth constraints. However, belief cascades may also derive from other possible mechanisms. In models of message sending with payoff interactions, message senders sometimes strategically report only coarsened versions of their beliefs (Crawford and Sobel (1982)). This raises the possibility that sequential reporting might result in cascades, which is an interesting topic for future research.[8]

naively by receivers as indicating that the state is L or H for sure. If, with some probability, receivers make this mistake, information cascades can start very quickly.

[8]Alternatively, in behavioral models with dual cognitive processing, an agent consists of two selves (sometimes called the “planner” and the “doer”) who face distinct decision problems. This can correspond, for example, to the System 2 versus System 1 thinking distinction of Kahneman (2011). Suppose that the planner rationally updates and assesses probabilities, and, when $\hat{p}_i > 0.5$ instructs the doer that the state is H, and, when $\hat{p}_i < 0.5$ instructs the doer that the state is L. Such simplified instructions may be needed owing to limited cognitive processing power of the doer, who may face problems of distraction and time constraints. If people are typically in “doer” mode when engaged in casual conversation with others, then only the coarsened information is reported. This would again result in cascades in reported beliefs. Furthermore, these reported beliefs correspond to the genuine beliefs of people when in doer mode.

A.5 Observation of Others Can Reduce Decision Quality when Information Acquisition is Costly and Observation of Others is Noisy

We now describe a setting in which social observation results in decisions that are even worse than the decisions that agents would make under informational autarky. This extends the discussion in the main text at the end of §7.2.

As mentioned in the main text, there is evidence suggesting that observation of others sometimes results in degradation in decision quality (a zoological example is provided by Gibson, Bradbury and Vehrencamp (1991)). In social psychology, "groupthink" (Janis and Mann (1977)) is the phenomenon that group deliberations sometimes result in disastrous decision failures, as if interaction with others were harming instead of improving decisions. Social psychologists have offered explanations for this phenomenon based upon imperfect rationality. However, even under rational social learning the observation of others can degrade decision quality.

This outcome may seem surprising, since any information an agent gleans by an agent via social observation is incremental to her own private information. However, this effect can be outweighed by the effect of informational free-riding, so that agents underinvest in private information, and instead mimic the actions of predecessors with noise.

To see why agents in groups can come to decisions that are on average worse than if there were no social observation, first consider the SBM, in which others are observed without noise, with the modification that there is a small cost of acquiring private signals. As discussed at the start of § 7.1, starting with I_2 all agents follow I_1, so the social belief reflects only a single signal. This is no more accurate than if agents decided independently (though welfare is higher as agents save on investigation costs).

Suppose instead that observation of predecessors is noisy, where each agent observes binary signals about the actions of all predecessors. Suppose further that all agents observe the same binary noisy signal about the action of any given predecessor.

If the noise is sufficiently small relative to the cost of the signal, the net gain to I_2 of acquiring a private signal is still negative, so she still does not do so. But now, owing to observation noise, her action is less accurate than if she were to decide on her own. So observation of others reduces decision quality relative to informational autarky. (Nevertheless, I_2's welfare is higher than under autarky, as observation of others economizes on observation costs.)

What about later agents? Agent I_3 also just follows I_3's signal about I_1's action.[9] The same applies to all later agents, so everyone's action is less accurate than if they had decided independently.

Decisions in groups can become even less accurate if agents observe only the latest predecessor. In this case, noise can compound repeatedly. However, a point is eventually reached at which an agent again pays to acquire a private signal (Cao and Hirshleifer (1997)), so there are cyclical shifts in accuracy.

[9]Agent I_3 ignores her signal about I_2's action, because she knows that I_2 imitated I_1 based on the same signal realization about I_1's action that I_3 observes. So if I_3's signal about I_2's action differs from I_3's signal about I_1's action, I_3 knows that this discrepancy *must* be caused by error in observation of I_2's action.

References

Cao, H. Henry, and David Hirshleifer. 1997. "Limited observability, reporting biases, and informational cascades." Working paper, Fisher College of Business, The Ohio State University.

Crawford, Vincent P., and Joel Sobel. 1982. "Strategic Information Transmission." *Econometrica*, 50(6): 1431–1451.

Daughety, Andrew F., and Jennifer F. Reinganum. 1999. "Stampede to Judgment: Persuasive Influence and Herding Behavior by Courts." *American Law and Economics Review*, 1(1 & 2).

Ellison, Glenn, and Drew Fudenberg. 1993. "Rules of thumb for social learning." *Journal of Political Economy*, 101(4): 612–643.

Ellison, Glenn, and Drew Fudenberg. 1995. "Word-of-Mouth Communication and Social Learning." *Quarterly Journal of Economics*, 110(1): 93–125.

Gibson, Robert M., Jack W. Bradbury, and Sandra L. Vehrencamp. 1991. "Mate Choice in Lekking Sage Grouse Revisited: The Roles of Vocal Display, Female Site Fidelity, and Copying." *Behavioral Ecology*, 2(2): 165–180.

Janis, Irving L., and Leon Mann. 1977. *Decision Making: A Psychological Analysis of Conflict, Choice, and Commitment.* New York:Free Press.

Kahneman, Daniel. 2011. *Thinking, Fast and Slow.* New York, NY:Farrar, Straus and Giroux.

Sunstein, Cass R. 2009. *A Constitution of Many Minds: Why the Founding Document Doesn't Mean What It Meant Before.* Princeton, NJ:Princeton University Press.

Talley, Eric. 1999. "Precedential Cascades: An Appraisal." *Southern California Law Review*, 73: 87–137.

Vermeule, Adrian. 2012. *Law and the Limits of Reason.* . Reprint ed., Oxford University Press.