# Exploitative informing

**Abstract**

Informing others about the world is typically regarded as a helpful act. In this paper, I study agents who conduct experiments to gather information about the world, committing in advance to fully disclose the nature of the experiment together with all experimental findings. While this appears to be a benign activity, I characterize a type of exploitative informing that is possible even within this restricted setup. I show how exploitative informants use public experiments to predictably manipulate interlocutors' beliefs and actions to their own advantage. I discuss epistemic and practical grounds on which it may be permissible for agents to resist acts of exploitative informing, then conclude by discussing implications for epistemic injustice and duties to gather evidence.

## 1 Introduction

Suppose you are a civil court judge.[1] Every week, BigFarm brings you a case seeking to repossess some SmallFarm or other. Both you and BigFarm's lawyer know no more about the case than what you have learned from past experience: BigFarm is 30% likely to be correct in cases of this sort.

As a civil court judge, you must find on the preponderance of evidence, so you won't find for BigFarm unless you are at least 50% confident that BigFarm owns the land. Just as you are about to laugh the case out of court, BigFarm's attorney proposes an experiment. The statistical properties of this experiment are known, and the results of the experiment will be accurately disclosed to you. This experiment has the following properties: in cases where BigFarm is entitled to the land, the experiment is certain to return a positive finding. In cases where BigFarm is not entitled to the land, the signal returns a positive finding with probability 3/7 and a negative finding with probability 4/7.

Surely, you think, there could be no harm in admitting this experiment. You do so and update your beliefs by conditionalization based on the observed result of the experiment, your prior knowledge, and the statistical properties of the experiment. A bit of algebra

[1]This case is adapted from Kamenica and Gentzkow (2011).

reveals an interesting pattern.[2] 40% of the time, the experiment will send a negative signal. You will then update to 0% credence in BigFarm's claim to the land and find against BigFarm. So far so good. But the other 60% of the time, the experiment will send a positive signal. You then update to 50% credence in BigFarm's claim to the land, and find for BigFarm.

Week after week, this happens, and 60% of the time, you award SmallFarm's land to BigFarm. This seems strange to you. After all, both you and BigFarm's attorney know that BigFarm is entitled to only 30% of these farms. Yet BigFarm is able to gain twice as many farms by conducting an honest, observable experiment week after week. In this case, there are two natural reactions. First, it looks like you are being exploited. BigFarm is predictably manipulating your beliefs to gain the verdicts they desire more often than they deserve them. And second, it seems permissible for you to resist BigFarm's exploitation. BigFarm is behaving badly and should be stopped. These are natural thoughts, but cashing them out is no mean feat.

Regarding exploitation, it seems odd to say that BigFarm's lawyer has exploited you by truthfully telling you all that they know about the case, then going out to gather relevant evidence and truthfully disclosing this evidence too. Nevertheless, in this paper I draw on recent discussions of Bayesian persuasion (Ely and Szydlowski 2020; Kamenica and Gentzkow 2011; Kamenica 2019) to characterize a species of exploitative informing (Section 2) and argue that BigFarm's testimony is a prime example. In particular, I study five ways in which many acts of exploitative informing are exploitative: they exploit asymmetric power to constrain deliberative outcomes (Section 3) in order to manipulate their interlocutor's acts (Section 4) and thereby monopolize the utility windfall from gathered information (Section 5). Exploitative informing worsens when it interacts with ongoing structures of domination (Section 6) and when it manipulates coarse-grained properties of an agent's doxastic state (Section 7). I argue that acts of information provision with many or most of these features have a good claim to be understood as exploitative.

[2] See Section 3 for details.

Regarding permissible resistance, it may seem odd to say that you may permissibly turn down the cost-free evidence provided by BigFarm's lawyer. Orthodoxy suggests that neither the expected utility of your decisions (Good 1966) nor the expected accuracy of your credences (Maher 1990) can decrease as a result of receiving cost-free evidence. In Section 8, I present the prima facie case against the permissibility of resisting exploitative informing. Then I sketch epistemic and practical grounds on which resistance may yet be permissible. Section 9 concludes by discussing implications of this discussion for our understanding of epistemic injustice (Section 9.1) and duties to gather evidence (Section 9.2), as well as some potential generalizations of this discussion (Section 9.3).

## 2 Exploitative informing

*Exploitative informing* occurs when an informant provides information to a receiver for the purpose of manipulating the receiver's actions or beliefs in order to achieve the informant's goals. It is perhaps not so surprising that exploitative informing should exist, or that receivers may permissibly resist exploitative informing. However, in this paper I will be concerned with cases of exploitative informing with four features that make both the possibility of exploitative informing and the permissibility of resisting it substantially more surprising.[3]

First, agents share a *common prior* $c$ over a common state-space $\Omega$. This means that the informant cannot exploit asymmetric possession of information to achieve her goals, since she plots her exploitation from the same epistemic position as the agent to be exploited.

Second, informants are *truthful*. Informants construct an experiment $\pi$ yielding valuable information about the world, then truthfully disclose the results of $\pi$. This means that the informant cannot exploit through deception.

Third, there is *no withholding* of information. Prior to running the experiment $\pi$, the informant commits to fully disclose the results of $\pi$, even if they are unfavorable. This

[3]These features are characteristic of recent discussion of Bayesian persuasion (Dworczak and Martini 2019; Kamenica 2019).

means that the informant cannot exploit through withholding unfavorable information any more than she can exploit by distorting unfavorable information.

Finally, the setup is *transparent* in the sense that all relevant parameters are common knowledge between the receiver and the informant. Each knows the prior $c$, both agent's utility functions, the space of possible experiments, the problem setup, and the constraints of truthfulness and non-withholding. This creates a remarkable situation in which receivers can be exploited in plain sight, with full prior knowledge that they are going to be exploited, who will exploit them, and why and how they are being exploited.

We will see in Section 8 that many theories of rationality say it is permissible, or even required for the receiver to accept exploitative informing. However, in this paper I argue that exploitative informing involves five types of manipulation that may permit or even require receivers to resist.

First, exploitative informing involves *asymmetric constraining power*. Experiments $\pi$ can be constructed to yield many probability distributions $\tau$ over resulting posteriors. I show in Section 3 that receivers exercise only a weak constraint over $\tau$, inherited from Bayes' Theorem. Informants then have the power to induce any distribution $\tau$ of their liking, so long as it respects Bayesian Constraint. This constitutes an asymmetry in power to determine the results of updating.

Second, exploitative informing involves *act manipulation* (Section 4). The informant constructs experiments to shift the receiver's probabilities of taking candidate actions, with an eye only to favoring actions that benefit the informant.

Third, exploitative informing involves *surplus hogging*. Information gathering typically generates a utility surplus, in the form of an improvement in the expected utilities of resulting actions. I show that exploitative informants construct experiments in order to reap the maximum possible personal surplus, leaving the receiver with no surplus whatsoever (Section 5). This hogging of the utility surplus is especially egregious because a less-selfish signal could generate a larger surplus to be shared between the two agents.

Fourth, exploitative informing gains bite when it *interacts with ongoing structures of* *domination*. Under diachronic relationships of exploitative informing, informants are able to reap larger surpluses than they could in one-off relationships, and those surpluses are further sharpened if informants have the power to determine the tasks that receivers perform (Section 6). These gains typically come at the direct expense of reduced benefits to interlocutors.

Fifth, exploitative informing *manipulates coarse-grained properties of belief* (Section 7). Insofar as it makes sense to think of the receiver's credences as encoding coarse-grained features such as false positive and false negative rates, informants have a robust ability to manipulate these coarse-grained features to their advantage. In large part, exploitative informants gain by providing information that does not reduce the frequency of coarse-grained error, but rather shifts the coarse-grained errors that are likely to be made in a direction maximally favorable to the informant.

We will also see that exploitative informing often provides *no epistemic benefit* along many traditional dimensions (Section 8.1). For example, exploitative informing may not improve the expected accuracy or calibration of the receiver's beliefs. While receivers may not lose along these dimensions, neither do they gain, and if they do gain their gain is often paltry and insignificant.

Summing up, exploitative informing occurs when informants provide information to receivers for the purpose of manipulating the agent's actions or beliefs in order to achieve the informant's goals. I am interested in cases involving a common prior, truthful disclosure without withholding, and a transparent problem setup. In these cases, I will argue that exploitative informing is problematically exploitative in many ways: it uses asymmetric constraining power to manipulate an agent's acts, hogging the surplus generated by new information. Exploitative informing is worsened by interaction with ongoing structures of domination, and at the epistemic level involves manipulation of coarse-grained properties of belief without benefit along traditional epistemic dimensions.

The next order of business is to show how exploitative informing involves these five forms of exploitation. I do this in Sections 3-7.

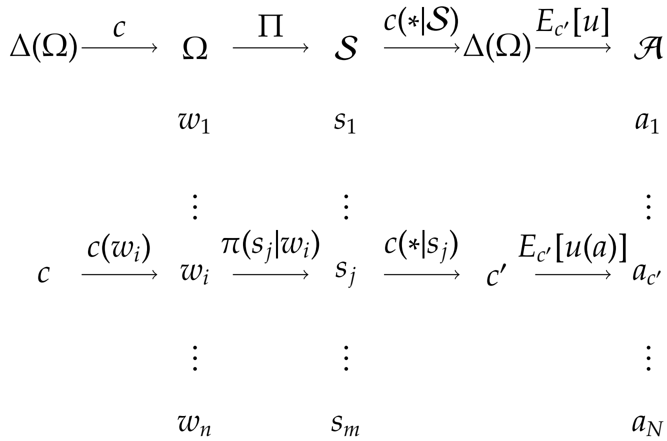


Figure 1: Choice under exploitative informing. Agents share a common prior $c$ over states $\Omega$. Sender $S$ performs an experiment $\pi$, leading to outcome $s_j$. Both agents update by Bayesian conditionalization to $c' = c(*|s_j)$. Receiver $R$ chooses the act $a_{c'}$ maximizing the expectation of her utility function $u$ given the posterior $c'$.

# 3 Asymmetric constraining power

To see how exploitative informing involves an asymmetry in the power to constrain outcomes, let us formalize the setup introduced in Section 1 (See Figure 1). There are two agents, Receiver $R$ and Sender $S$. In this case, $R$ is the judge and $S$ the BigFarm lawyer. Both agents share a common prior $c$ over the algebra generated by a finite state space $\Omega$, letting $\Delta(\Omega)$ represent the set of possible priors over this algebra.

Receiver $R$ must choose among a finite set $\mathcal{A}$ of acts. In this case, she must choose between the act $F$ of finding for SmallFarm and $B$ of finding for BigFarm. Agents $R, S$ receive utilities $u(a(w)), v(a(w))$ dependent on the outcome $a(w)$ of the chosen act $a$ in the current state $w$.

In this case, there are two relevant states: the state $w_b$ in which BigFarm is entitled to the land, and the state $w_f$ in which SmallFarm is entitled to the land. Our judge $R$ gains from just decisions, whereas the lawyer $S$ gains from victory, so that $u, v$ are given by Figure 3. $R$ may choose whether or not to perform an experiment $\pi$ to gather information about the world. Experiments are specified by the set $\mathcal{S}$ of possible outcomes, and the state-dependent probabilities $\pi(s_j|w_i)$ of each outcome $s_j \in \mathcal{S}$ in each world-state $w_i \in \Omega$.

| | u | | | v | |
|---|---|---|---|---|---|
| | B | F | | B | F |
| $w_b$ | 1 | 0 | $w_b$ | 1 | 0 |
| $w_f$ | 0 | 1 | $w_f$ | 1 | 0 |

Figure 2: Payoffs to Receiver ($u$) and Sender ($v$) contingent on states ($w_b, w_f$) and acts ($B, F$).

By assumption, the statistical properties $\pi$ and possible outcomes $\mathcal{S}$ of experiments are known by both parties, and $R$ is bound to truthfully report the outcome of any experiment she performs.

In our example, BigFarm performs an experiment $\pi$ with two possible outcomes, $s_1, s_2$, with the intended interpretation that $s_1, s_2$ are positive and negative findings, respectively. In $w_b$, BigFarm is entitled to the land and $\pi$ verifies BigFarm's ownership with certainty, so that $\pi(s_1|w_b) = 1$ and $\pi(s_2|w_b) = 0$. In $w_f$, BigFarm isn't entitled to the land, but the experiment finds positively 3/7 of the time, so that $\pi(s_1|w_f) = 3/7$ and $\pi(s_2|w_f) = 4/7$.

If $R$ performs no experiment, $S$ takes the act $a_c$ maximizing expected utility given her prior $c$ and utility function $u$.[4] In our example, $R$ finds for SmallFarm. However, if $R$ performs an experiment $\pi$, then $R$ updates her beliefs $c$ on the resulting signal $s$, to $c' = c(*|s)$. $R$ then takes the act $a_{c'}$ maximizing expected utility given her updated beliefs $c'$. In our example, a little algebra shows that signal $s_1$ is 60% likely to be sent, in which case $c'(w_b) = 0.5$, and signal $s_2$ is 40% likely to be sent, in which case $c'(w_b) = 0$. When $c'(w_b) = 0.5$, the judge finds for BigFarm, whereas when $c'(w_b) = 0$, the judge finds for SmallFarm.[5]

Note that the experiment $\pi$ induces a probability distribution $\tau$ over possible posteriors $c'$, due to uncertainty over the signal that will be received (Figure 3). In this case, $\tau$ gives 60% probability to the case where $c'(w_b) = 0.5$, and 40% probability to the case where $c'(w_b) = 0$. Receiver has the power to impose one, and only one constraint on $\tau$: the

[4]We can absorb this case into the setup described in Figure 1 by regarding $R$ as performing the trivial experiment in which $\pi$ is the identity function.

[5]Signal $s_2$ is sent 4/7 of the time in state $w_f$, which has a 70% chance of occurring, so signal $s_2$ is sent 40% of the time and signal $s_1$ is sent the other 60% of the time. Signal $s_2$ can only be sent in $w_f$, so $c(w_b|s_2) = 0$. However, $c(w_b|s_1) = c(s_1|w_b)c(w_b)/c(s_1) = \frac{c(s_1|w_b)c(w_b)}{c(s_1|w_b)c(w_b)+c(s_2|w_f)c(w_f)} = \frac{(1)(3/10)}{(1)(3/10)+(3/7)(7/10)} = 1/2$.

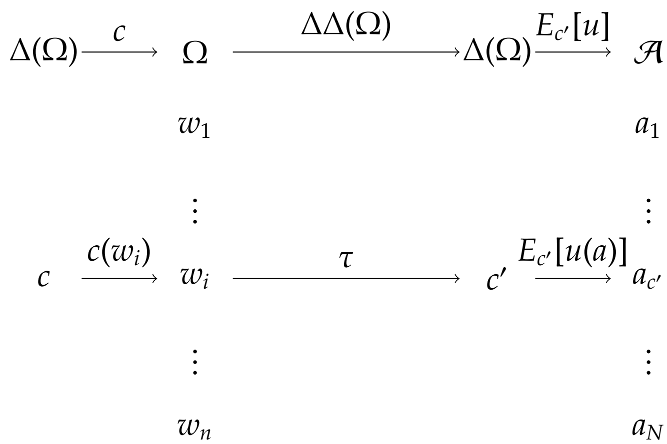


Figure 3: The consequences of experiments are captured by the probability distributions $\tau$ they induce over shared posteriors $c'$.

induced distribution over posteriors must be consistent with Bayes' Theorem. That is:

**(Bayesian Constraint)** $E_\tau(c') = \sum_{c'} \tau(c')c' = c$.

However, Bayesian Constraint is a slim constraint. For example, Figure 4 gives a small sample of the many different distributions $\tau$ consistent with Bayesian Constraint. Does Receiver have any constraining power beyond Bayesian Constraint?

It turns out that she does not. Any distribution $\tau$ over posteriors consistent with Bayesian Constraint can be induced by some experiment $\pi$ (Kamenica 2019). In this sense, Sender $S$ enjoys highly asymmetric power to constrain outcomes. Receiver $R$ can only impose the standard Bayesian Constraint, which any Bayesian agent is guaranteed to satisfy. However, short of refusing to learn or conditionalize on the outcome of Sender's experiment, Receiver cannot impose any further constraints on the distribution of posterior credences, no matter her goals or motivations. This leaves Sender free to choose any of the vast number of distributions $\tau$ over posteriors consistent with Bayesian Constraint. For example, Sender can choose to impose any one of the distributions $\tau$ in Figure 4. Sender is free to choose the distribution $\tau$ which benefits her the most, bound only, if at all, by the contingent availability of resources and technology to perform experiments left open by Bayesian Constraint.

| $c'(w_b)$ | $\tau(c')$ |
|---|---|
| 0 | 0.7 |
| 1 | 0.3 |

| $c'(w_b)$ | $\tau(c')$ |
|---|---|
| 0.2 | 0.875 |
| 1 | 0.125 |

| $c'(w_b)$ | $\tau(c')$ |
|---|---|
| 0 | 0.5 |
| 0.5 | 0.4 |
| 1 | 0.1 |

Figure 4: Distributions $\tau$ over common posteriors $c'$ consistent with Bayesian Constraint.

# 4 Act manipulation

How does Sender $S$ manipulate Receiver $R$? Since $S$'s payoffs depend on $R$'s acts, there is only one way for $S$ to benefit from manipulation: through *act manipulation*: $S$ changes the probabilities that $R$ will take certain actions.

We saw in Section 3 that an experiment $\pi$ directly induces a probability distribution $\tau$ over posteriors. Each posterior determines $R$'s act together with her utility function, hence $\pi$ *indirectly* induces a distribution $\tau_{\mathcal{A}}$ over acts that $\mathcal{A}$ will take. $S$ picks an experiment to maximize her utility from $\tau_{\mathcal{A}}$, namely $\sum_{a \in \mathcal{A}} \tau_{\mathcal{A}}(a) E_c[v(a)]$.

In our example, $R$ begins with a certain disposition to $F$, finding for SmallFarm. She ends up 60% likely to find for BigFarm, and only 40% likely to find for SmallFarm. This is surprising for two reasons. First, $R$ is quite confident that BigFarm is correct in only 30% of similar cases. However, if BigFarm conducts the same experiment in this and similar cases, then over the long run, BigFarm will win 60% of the time. Judge $R$ may feel herself oddly helpless as, time and time again, she finds for BigFarm despite knowing that BigFarm is winning twice as many cases as they are entitled to win.

A second oddity is that Sender has pushed the judge's act probabilities $\tau_{\mathcal{A}}$ as far as they can go, while still respecting Bayesian Constraint. Bayesian Constraint says that 30% probability mass must be put on $B$, the claim that BigFarm is entitled to the land. BigFarm's lawyer knows that the judge will only find for BigFarm when $c'(w_b) \geq 0.5$. This can happen a maximum of 60% of the time, namely if $c'(w_b)$ is exactly equal to 0.5 in 60% of the cases and equal to 0 the rest of the time. Hence BigFarm's lawyer induces a distribution $\tau$ over posteriors with exactly this feature, so that the resulting distribution

$\tau_{\mathcal{A}}$ over acts will maximize the probability of Receiver acting in BigFarm's interest.

This result is also fully general: insofar as she is concerned only with her own benefit, Sender's optimal strategy is *always* to induce the most favorable act distribution $\tau_A$ that can be reached while respecting Bayesian Constraint. That is, she manipulates Receiver's act dispositions to be as favorable as possible, limited only by Bayesian Constraint.

In light of these results, it may be helpful to view act manipulation as the practical analog of asymmetric constraining power. We saw in Section 3 that Sender is free to impose any distribution $\tau$ over posteriors, limited only by Bayesian Constraint. Since acts are determined by Receiver's fixed utility function, together with the shared posterior, this means that Sender is also free to impose any probability distribution $\tau_A$ over resulting acts, limited only by Bayesian Constraint. In this sense, Sender enjoys asymmetric power to constrain the beliefs and acts that will result from experimentation. In both cases, Sender exercises her constraining power to choose the most favorable distributions $\tau$ over posteriors and $\tau_A$ over resulting acts.

# 5 Surplus hogging

In its classic sense, the expected value of information to an agent is determined by the expected change in the value of the acts that will be taken after becoming informed (Good 1966; Howard 1966). Before receiving information, Receiver is certain to take the best act $a_c$, with expected value $E_c[u(a_c)]$. After conducting experiment $\pi$, she adopts some (uncertain) posterior $c'$ with probability $c(c'|\pi)$ incorporating the prior together with the statistical properties of experiment $\pi$. She then takes the act $a_{c'}$ maximizing expected utility on $c'$, and receives expected utility $E_{c'}[u(a_{c'})]$.

For Receiver, the value of information $\pi$ is the expected change in the value of her acts, namely $VOI(\pi, R) = \sum_{c'} c(c'|\pi)E_{c'}[u(a_{c'})] - E_c[u(a_c)]$. This quantity is guaranteed to be nonnegative by Good's Theorem (Good 1966). In this sense, we will see in Section 8.2 that many have thought Receiver cannot rationally prefer to avoid receiving exploitative

| | Expected prior utility | Expected posterior utility |
|---|---|---|
| $R$ | 0.7 | 0.7 |
| $S$ | 0 | 0.6 |
| Total | 0.7 | 1.3 |

Figure 5: Surplus hogging

information.

The value of information to Sender is identical, except that she is rewarded according to her own utilities $v$, so that $VOI(\pi, S) = \sum_{c'} c(c'|\pi)E_{c'}[v(a_{c'})] - E_c[v(a_c)]$. This can be negative for some experiments $\pi$, but won't be negative for any experiments chosen by Sender. After all, Sender can always conduct an uninformative experiment and reap no gain or loss, so she won't choose to conduct any experiment with negative expected value of information.

So far, we have seen that the value of information from chosen experiments is certain to be nonnegative for all agents. However, typically the value of information is not merely nonnegative, but strictly positive. Gaining information helps agents to make better decisions. In this case, a *utility surplus* is reaped, equal to $VOI(\pi, R) + VOI(\pi, S)$.

In our example, Receiver is previously disposed to find for SmallFarm, providing her with utility 0.7 and BigFarm with utility 0 (Figure 5). Now, Receiver is disposed to find for BigFarm 60% of the time, though she will be wrong in half of those decisions, providing her with utility 0.7 and BigFarm with utility 0.6. There is, in this example, a utility surplus of 0.6. But that surplus is reaped entirely by BigFarm. Receiver gains nothing from BigFarm's experiment except the opportunity to help BigFarm gobble up SmallFarm.

More generally, exploitative informants seek to maximize their own surplus rather than the total surplus value of information. In this sense, Sender is a *surplus hog*: she keeps as much as possible of the surplus for herself.

An analogy may reveal why many find surplus hogging to be objectionable. Consider:

**Dictator Game:** Dictator is told to propose a division of $10 between herself and Peon. Peon chooses whether to accept or reject the proposed division. If Peon accepts, the money is divided as proposed. If Peon does not accept, no player receives any money.

Classic economic theory proposes that Dictator will propose to hog the surplus $10: she might, for example, offer Peon a nickel, keeping $9.95 for herself. But in fact, that is often not what happens, and with good reason: when Peon is offered only a small share of the total, she tends to reject the proposed division, leaving both players with nothing (Engel 2011).

Why might Peon turn down a free nickel? One popular answer is that Peon thinks Dictator's hogging of the surplus is exploitative and unfair (Hoffman et al. 1996; Engel 2011). To her mind, giving up a free nickel, or even as many as two or three free dollars is a small price to pay for stopping Dictator's exploitation.

We might understand Receiver in our game as experiencing a similar reaction. What Sender proposes to do is to conduct an experiment that generates a utility surplus, which she will then hog for her uses, giving none of the surplus to Receiver. Just as Peon may resent the exploitative division of surplus money proposed by Dictator, so too Receiver may resent the exploitative division of surplus value from information proposed by Sender.

In fact, our current problem is worse than the Dictator Game, for the surplus value of information is not fixed. In many cases, Receiver could *increase* the utility surplus generated by her experiment by choosing an experiment that distributes the surplus more equitably among parties.

For example, suppose that BigFarm gains utility 0.5 in cases where the judge finds for SmallFarm, perhaps because BigFarm has taken out limited insurance on the case. Now, the utilities are given by Figure 6. In this case, BigFarm's optimal experiment is the same as before: she performs an experiment which is certain to send a positive signal in case $w_b$, she is entitled to the land, and which also sends a positive signal 3/7 of the time in $w_f$, where she is not entitled to the land. Calculating as before shows that BigFarm reaps expected posterior utility 0.8 and the judge reaps expected posterior utility 0.7, for a total of 1.5.

u

| | B | F |
|---|---|---|
| $w_b$ | 1 | 0 |
| $w_f$ | 0 | 1 |

v

| | B | F |
|---|---|---|
| $w_b$ | 1 | 0.5 |
| $w_f$ | 1 | 0.5 |

Figure 6: Modified utilities leading to surplus diminution.

However, if BigFarm were not a selfish rascal, she could increase the total posterior utility split between her and the judge by conducting a different experiment. Suppose that BigFarm conducts an experiment $\pi'$ fully disclosing the state of the world. The judge then reaps expected posterior utility 1, since she is certain to find justly. BigFarm reaps expected posterior utility 0.65, since she is now only 30% likely to get the farm. But the total expected posterior utility has increased to 1.65.

Such cases reveal the full depth of surplus hogging. Exploitative informants act with an eye to maximizing the expected value that informational disclosures will provide to themselves. This is not merely a zero-sum game of stealing posterior utility away from others. In many cases, exploitative informing is a negative-sum activity, driving down the total utility gained from informational disclosures. In this sense, exploitative informants are perhaps worse than dictators who propose an uneven division of a ten-dollar prize. Exploitative informants are often better likened to a dictator who proposes to pay herself six dollars instead of a scenario in which she and another agent each receive five dollars. This type of surplus hogging may be objectionable not merely on grounds of inequity, but also on the grounds that the surplus itself has been shrunk.

# 6 Interaction with ongoing structures of domination

Exploitative informing is worsened as it interacts with ongoing structures of domination, advantage and control (Ely 2017; Ely and Szydlowski 2020). Let us begin by considering what happens if BigFarm's asymmetric ability to conduct experiments is extended over time.

As it happens, BigFarm's lawyer $R'$ is repeatedly hired by BigFarm, $S'$, to sue small farms. It's hard to muster much sympathy for the farm-stealing rascal, but actually $R'$ is not in such a good way herself. In each case, BigFarm's lawyer $R'$ must invest time $\tau$, with time investment $\tau = 0$ standing in for declining the contract. Some cases are easy, requiring 50 hours to prosecute effectively, whereas others are easy, requiring 100 hours to prosecute effectively. BigFarm has never been fond of looking before it litigates, so neither BigFarm nor BigFarm's lawyer knows whether any given case is hard or easy, though they do have a common prior $c$ over case difficulties gleaned from years of stealing farms.

Lawyer $R'$ is paid a salary worth utility $P$ for prosecuting cases effectively, but she is paid nothing for ineffective prosecution. Letting $x \in \{50, 100\}$ be the unknown number of hours required to effectively prosecute the current case, and $\tau$ be the number of hours she invests in prosecuting the case, lawyer $R'$ is paid $P$ if $\tau \geq x$ and nothing otherwise. Let us assume that time investments cost $R'$ one util per hour in missed opportunities. Then from any given time investment $\tau$, lawyer $R'$ reaps expected utility:

$$u(\tau) = c(\tau \geq x)P - \tau.$$

To begin, suppose that $R'$ is 80% confident the task is easy, requiring 50 hours to receive $P$, and 20% confident the task is hard, requiring 100 hours to receive $P$. Set the reward $P$ at 75 utils.

In this case, $R'$ is disposed to accept the case and invest 50 hours in prosecuting it. As a result, she receives an expected payoff of $0.8 * 75 - 50$, or 10 utils.

Now, suppose that BigFarm desires only to maximize the probability that her lawyer prosecutes the case effectively. BigFarm has the ability to conduct an experiment $\pi$ informing her lawyer $R'$ about the difficulty of the current task. Can BigFarm gain from experimentation?

It is easy to see that in the present case, BigFarm does not gain from experimentation. No experiment could ever convince the lawyer to invest 100 hours in the case, since this represents a certain loss of 25 utils. Nor will she ever invest strictly between 50 and 100 hours in the case, since this costly effort is certain to yield no improvement in her reward over an investment of 50 hours. The most that our lawyer $R'$ will invest after experimentation is 50 hours. Since $R'$ is already disposed to invest 50 hours, the only thing that an informative experiment could do is to convince $R'$ to run away and invest 0 hours, by making her highly confident that the case is hard. However, suppose that BigFarm desires to extract more effort from her lawyer and also to reduce the lawyer's utility surplus to zero. Surprisingly, she can do both of these things if we examine how the setup interacts with ongoing diachronic structures of domination inherent in a contractive labor relationship.

Let us consider a diachronic version of the same case. Now BigFarm has the power to precommit to conducting an experiment $\pi$ after 50 hours have been invested in the case. Experiment $\pi$ will help lawyer $R'$ to understand whether more effort is needed to reap her reward. BigFarm conducts an experiment that fully discloses the task difficulty. That is, the experiment has two outcomes $s_1, s_2$, with $\pi(s_1|x = 100) = \pi(s_2|x = 50) = 1$.

If $\pi$ reveals that the task is easy, lawyer $R'$ halts and collects her reward. However, if $\pi$ reveals that the task is hard, $R'$ is now disposed to continue. After all, she has already sunk 50 hours of labor into the case, so the expected payoff of 50 additional hours is $P - 50 = 25$ utils. Here, BigFarm's communicative power gives her the ability to ensure with certainty that her lawyer invests maximum effort into the case, even though the lawyer was previously disposed to invest moderate effort into the case, and no synchronic signal could induce her to do otherwise.

Although BigFarm gains power to extract labor from her lawyer, we should not yet feel bad for $R'$. After all, $R'$ has become richer. She expects that 80% of the time, the signal will reveal the case to be easy, in which case she halts and reaps total utility $75 - 50 = 25$. She expects that the remaining 20% of the time, the signal will reveal the case to be hard, in which case she works a total of 100 hours and reaps a loss of $75 - 100 = -25$ utils. In expectation, she now receives 15 utils, an improvement in her expected payoff of 10 utils from the synchronic problem.

Unfortunately, I neglected to mention that BigFarm, as a corporation, has the power to set the salary $P$ it will pay for effective prosecutions. After consulting with the in-house philosopher, BigFarm reduces its standard salary promise $P$ to a value of 60 utils. What are we to make of this revised situation?

BigFarm could not get away with this salary change if it did not have the power to conduct informative experiments. After all, with no additional information, even 50 hours of effort now represents an expected loss of 2 utils. However, BigFarm does have the ability to convey information after 50 hours, and that power makes all the difference.

Now, BigFarm promises $R'$ that she will fully disclose task difficulty after 50 hours. This is just enough to convince $R'$ to take the case. After all, 80% of the time, $R'$ expects to learn that she is done after 50 hours, and will earn a net profit of 10 utils. The other 20% of the time, $R'$ expects to learn that she has made a loss of 50 utils, which can be reduced to 40 utils after slaving away for another 50 hours. In net, $R'$ expects neither to gain nor to lose from investing 50 hours of work into the case.

As before, information does provide a surplus, but that surplus is entirely reaped by BigFarm. Her lawyer went from a certain disposition to decline the case, with expected utility zero, to a certain disposition to prosecute the case fully, still with expected utility zero. BigFarm, by contrast, moves from her least-preferred situation in which the case is never prosecuted into her most-preferred situation in which the case is always prosecuted fully.

In fact, that is not the end of it. Recent work has shown that if the reward structure of the problem is still not enough to induce BigFarm's lawyer to take the case, then she can often divide the case into subproblems, each with a communicative experiment at the end that is just sufficient to induce the lawyer to perform the subproblem (Ely and Szydlowski 2020). In the present case, such manipulations would allow BigFarm to reduce her salary offer below \$60. But I think we have seen enough. Ongoing structures of domination, involving a diachronic asymmetry in the ability to conduct informative experiments, allowing dominating firms to extract more effort through informational disclosures from those they dominate. If these structures are strengthened to involve other powers, such as price-setting, the surplus can be entirely passed to the firm, leaving even our farm-stealing lawyer with nothing to show for her additional efforts, which are the maximum possible efforts desired by the firm. Here exploitative informing directly harms BigFarm's lawyer, who would otherwise be able to demand higher wages for her services.

# 7 Coarse-grained properties

Sometimes, it makes sense to think of coarse-grained properties as supervening on fine-grained properties of credence. Most obviously, this is the case on a Lockean view where agents believe all and only claims above a probability threshold (Dorst 2019; Foley 2009). For example, we might think of our judge $R$ in the original case as a Lockean with belief threshold 0.5. Alternatively, we may think of many non-Lockean agents as having judgment or decision-thresholds fixed by their credences. For example, we might say that the judge judges that an agent deserves the farm just in case she is at least 50% confident that the agent deserves the farm, or that the judge treats 50% confidence as a sufficient threshold for the decision to award the farm to a given agent.

With respect to each of these coarse-grained notions, we can define four traditional statistical categories: true positive, true negative, false positive, and false negative. For example, the judge reaches a true positive when she finds that $B$, BigFarm is entitled to the farm, in the state $w_b$ where BigFarm is in fact entitled to the farm. The judge reaches a false positive when she finds $B$ in $w_f$. She reaches a true negative when she finds $\neg B$ in $w_f$ and a false negative when she finds $\neg B$ in $w_b$.

With respect to these four statistical categories, we can define a number of coarse-grained properties that agents may have reason to care about. In Figure 7, I have grouped these properties into three clusters of four, deriving the underlying list of relevant coarse-grained properties from a recent survey (Verma and Rubin 2018) so that I may not be

| | Definition | Prior | Posterior |
|---|---|---|---|
| **TP** | | 0 | 30 |
| **TN** | | 70 | 40 |
| **FP** | | 0 | 30 |
| **FN** | | 30 | 0 |
| **TPR** | $TP/(TP+FN)$ | 0 | 1 |
| **TNR** | $TN/(FP+TN)$ | 1 | 4/7 |
| **FPR** | $FP/(FP+TN)$ | 0 | 3/7 |
| **FNR** | $FN/(TP+FN)$ | 1 | 0 |
| **PPV** | $TP/(TP+FP)$ | 0 | 1/2 |
| **FDR** | $FP/(TP+FP)$ | 0 | 1/2 |
| **NPV** | $TN/(TN+FN)$ | 7/10 | 1 |
| **FOR** | $FN/(TN+FN)$ | 3/10 | 0 |

Figure 7: Coarse-grained attitudinal properties

accused of statistical cherry-picking. Properties are normalized to their average values among 100 identical cases.

In the first cluster, I report the four basic properties: true positive (TP), true negative (TN), false positive (FP), and false negative (FP). In Figure 7, we see that the lawyer's persuasion has the effect of converting negative verdicts to positive verdicts with indiscriminate abandon. A total of 60 negative verdicts in every 100 are shifted to positive verdicts in the posterior, split evenly between true and false positives. In this sense, persuasion does not reduce the probability that errors will be made, but simply shifts the types of errors that will be made: the lawyer has successfully swapped 30 false negatives for 30 false positives.

In the second cluster, I report one way of analyzing the rates at which the four basic properties occur. The true positive rate $TPR = TP/(TP + FN)$ gives the probability that a truly positive case will be recognized as such. The true negative rate $TNR = TN/(FP+TN)$ gives the probability that a truly negative case will be recognized as such. The false positive rate $FPR = FP/(FP + TN)$ gives the probability that a negative case will be incorrectly treated as positive. The false negative rate $FNR = FN/(TP + FN)$ gives the probability that a positive case will be incorrectly treated as negative. In Figure 7, we see that the true positive rate has been increased from 0 to 1, inducing a corresponding decrease in the false negative rate from 1 to 0. In this sense, the lawyer purchases true positive findings by reducing the rate of false negative findings. That might not sound like a bad thing until we realize how it is done. The true negative rate of 1 has been reduced to 4/7, allowing the false positive rate to rise from 0 to 3/7. The increased true positive rate is thus purchased at the expense of a 3/7 increase in the rate of false positives, and a 3/7 decrease in the rate of true negatives.

In the final cluster, I report another way of analyzing the rates at which the four basic properties occur. The positive predictive value $PPV = TP/(TP + FP)$ gives the probability that positive predictions are correct. The false discovery rate $FDR = FP/(TP + FP)$ gives the probability that positive predictions are incorrect. The negative predictive value $NPV = TN/(TN + FN)$ gives the probability that negative predictions are correct. The false omission rate $FOR = FN/(TN + FN)$ gives the probability that negative predictions are incorrect. In Figure 7, we see that the positive and negative predictive values of the judge's findings have gone up: the judge is now more likely to treat positive cases as positive and to treat negative cases as negative. But this increase in predictive value is purchased at advantage to BigFarm: half of all farms awarded to BigFarm are farms it does not own, whereas previously no such farms were awarded to BigFarm.

Together, these coarse-grained properties of the judge's beliefs, judgments or decision thresholds convey a strong impression that the judge's beliefs and resulting act dispositions have been manipulated to the advantage of BigFarm. BigFarm's lawyer has successfully swapped false positives for false negatives, purchasing an increased true positive rate at the expense of a 3/7 increase in the rate of false positives, and a 3/7 decrease in the rate of true negatives, and purchasing an increase in predictive value at the expense of a situation in which half of all farms awarded to BigFarm are farms it does not own. Coarse-grained properties of the judge's beliefs, judgments or decision thresholds have been manipulated here to the maximal advantage of BigFarm.

# 8 Resisting exploitative informing

So far, we have been concerned with cases of exploitative informing in which informants provide information to receivers for the purpose of manipulating the receiver's actions or beliefs in order to achieve the informant's goals. Surprisingly, we saw that exploitative informing is possible when agents share a common prior, informants are truthful and do not withhold information, and the setup is fully transparent.

It may seem perverse to suggest that informing could ever be exploitative, particularly under such open conditions. However, we saw that exploitative informing looks to be exploitative in at least five ways: it draws on asymmetric power to constrain the outcomes of deliberation (Section 3) in order to manipulate the receiver's acts (Section 4), monopolizes the surplus value of information for the informant's own use (Section 5), interacts with ongoing structures of domination (Section 6), and manipulates coarse-grained properties of the receiver's beliefs (Section 7). This suggests that it may be permissible for agents to resist exploitative informing, but saying why it is permissible to resist exploitative informing proves surprisingly difficult. In this section, I discuss epistemic and practical grounds on which agents might resist exploitative informing.

## 8.1 Epistemic grounds for resistance

The reason why it is difficult to find epistemic grounds for resisting exploitative informing is encapsulated in an epistemic extension of Good's Theorem due to Patrick Maher (1990). Suppose that agents score the accuracy of their credences using a rule that is strictly proper, which is to say that under the rule, every credence function is self-recommending. For example, she might use the Brier Score of assessing credences by their squared distance from the truth (Brier 1950; Joyce 1998). Maher proved that under these conditions, with a few additional assumptions, agents cannot rationally expect evidence-gathering to decrease the accuracy of their credences.

Maher's result applies equally to acts of exploitative informing: the receiver can gather evidence by learning the result of the experiment $\pi$. Insofar as she evaluates accuracy using a strictly proper scoring rule, she cannot expect the accuracy of her credences to be reduced by this act. In fact, the accuracy of her credences may well increase. Indeed, both the judge and the lawyer considered in this paper experience slight gains in the Brier accuracy of their beliefs as a result of exploitative informing. This means that insofar as agents are concerned to maximize the accuracy of their credences, it will be hard to find rational grounds for resisting exploitative informing short of a technical challenge to conditions of Maher's result. However, we might find grounds for resisting exploitative informing by broadening our conception of rational agents' goals.

First, note that Maher's result only concerns the *accuracy* of an agent's credences. Perhaps rational agents may care about other properties of their doxastic states. For example, an agent's credences are *well-calibrated* to the extent that, for each $X$, when she is $X$% confident in a proposition, she tends to be right $X$% of the time. Calibration gives the judge no grounds to accept manipulative informing, since she is already perfectly calibrated: she knows that 30% of farms belong to BigFarm, so she always has 30% confidence that a given farm belongs to BigFarm, and she is right 30% of the time. If rational agents are concerned with the calibration of their beliefs (Van Fraassen 1983; Shimony 1988), they may thus not have reason to comply with exploitative informing.

Similarly, we saw in Section 7 that in many cases, there are well-defined coarse grained properties of an agent's doxastic state, broadly understood. We saw that there is a clear sense in which exploitative informing does not make these properties any better, but rather shifts the balance of error in ways favorable to the informant. Receivers concerned with improving the coarse-grained properties of their doxastic state may have no grounds to accept exploitative informing. Some would suggest that when fairness or bias are at issue, receivers may even be permitted or required to resist (Basu 2019; Fazelpour and Danks 2021).

Second, we will see in Section 9.1 that exploitative informants may perpetrate a distinctively epistemic type of injustice, harming receivers in their capacity as a knower by manipulating receivers' beliefs for the sender's own ends. If this is right, then it may provide a distinctively epistemic ground for resisting exploitative informing.

Finally, even if there are no good epistemic grounds for resisting exploitative informing, exploitative informing often relies on acts which are not purely epistemic. For example, in the diachronic case (Section 6) agents perform fifty hours of labor before information is disclosed. The decision to perform fifty hours of labor can and must be made not only on epistemic grounds, but also by considering the overall utility of performing fifty hours of labor. And even in the synchronic case, the possibility of exploitative informing is a result of prior acts that constrain the availability of experiments and the permissibility of conducting and disclosing them. These acts may also be evaluable on practical grounds. This suggests that we may do well to look for practical grounds to resist exploitative informing.

### 8.2 Practical grounds for resistance

Just as the epistemic case for resisting exploitative informing is complicated by Maher's result, the practical case for resistance is complicated by Good's Theorem (Good 1966). Good proved that under several assumptions, agents cannot rationally expect the utility of their decision problem to decrease as a result of gathering information. The expected value of information may well be zero, due to surplus hogging, but it cannot be negative. This suggests that it may not be rational for receivers to strictly prefer to avoid exploitative informing.

One way to resist would be to lean on recent challenges to Good's Theorem, which argue that agents may assign negative expected value to information if they are risk averse (Buchak 2010; Campbell-Moore and Salow 2020), uncertain they will conditionalize (Neth forthcoming), or face a number of other circumstances (Adams and Rosenkrantz 1980; Bradley and Steele 2016; Kadane et al. 2008). This strategy will stand or fall with the strength of these challenges to Good's Theorem and their applicability to individual cases of exploitative informing.

Another way to resist leans on the discussion in this paper. We saw that exploitative informing involves asymmetric constraining power (Section 3) used to manipulate agents' acts (Section 4) in order to hog the surplus from gathered information (Section 5) in a way that interacts with ongoing structures of domination (Section 6). We might well build each of these features into our description of outcomes, thereby rendering them utility-relevant. For example, we could say that the outcome of exploitative informing involves unfairness, manipulation, domination or control, and that agents may rationally prefer to avoid outcomes with these features.

This move is familiar from recent discussions of consequentializing, which aim to widen the scope of normative phenomena that can be incorporated within broadly consequentialist and decision-theoretic approaches, often by expanding the scope of what is understood by the outcome of an action (Dreier 2011; Portmore 2011; Seidel 2019). Some think that consequentializing is nearly always successful, and that it shows how a wide variety of normative phenomena can be brought within the consequentialist fold. These theorists should be friendly to a broadening of how the outcomes of exploitative informing are described. Others may think that many instances of consequentializing are trivializing, unmotivated, or deprive consequentialist and decision-theoretic approaches of their explanatory power (Muñoz 2021; Schroeder 2007; Tenenbaum 2014). These theorists evaluate consequentializing maneuvers on a case-by-case basis, and they are welcome to apply a similar scrutiny to the broadening of outcomes suggested above.

Those unsatisfied with the consequentializing approach should probably see the permissibility of resisting exploitative informing as a datum located within nonconsequentialist normative theories. Many such approaches allow us to directly say that exploitative informing is wrong because it is manipulative, controlling, selfish, and interacts with structures of domination. They also allow us to say, on these grounds, that agents should resist exploitative informing. On nonconsequentialist approaches, if Good is correct that the expected utility of exploitative informing for the receiver cannot be negative, then this just shows that the permissibility of acts does not always supervene on their expected utilities.

Summing up, we have seen that there may be epistemic grounds for resisting exploitative informing. Although exploitative informing cannot reduce the expected accuracy of receivers' beliefs, exploitative informing may fare less well when we move beyond accuracy as an epistemic standard. We have also seen that there is room to mount a practical objection to many of the acts involved in exploitative informing. This challenge can be mounted within a consequentialist theory, in an attempt to make the challenge consistent with Good's Theorem, or it can be mounted within a nonconsequentialist framework without the need to reconcile itself with Good's Theorem.

# 9 Discussion

This section concludes by discussing three avenues for future research: connections between exploitative informing and epistemic injustice (Section 9.1), new grounds for duties to gather evidence (Section 9.2), and generalizations of the present discussion (Section 9.3).

## 9.1 Epistemic injustice

It is natural to say two things about acts of exploitative informing. First, many such acts are not merely exploitative but also unjust. These acts are unjust because they involve exploiting receivers in order to achieve senders' ends, depriving receivers of the benefits they could have received through information sharing. And second, though some of the injustice done through exploitative informing is practical, a substantial part of this injustice appears to be epistemic, harming the agent in her capacity as a knower. In paradigmatic cases of exploitative informing, senders manipulate receivers' beliefs in order to use the receiver's beliefs as means to the sender's own ends. Combining these thoughts suggests that there is a kind of epistemic injustice involved in many acts of exploitative informing. If this is right, then three conclusions suggest themselves.

First, it would be productive to theorize directly about the type of epistemic injustice involved in exploitative informing, and in particular to relate this injustice to existing taxonomies and theories (Dotson 2014; Medina 2017; Fricker 2007, 2017). Second, we saw in Section 8.2 that one way to resist the argument from Good's Theorem to the impermissibility of resisting exploitative informing would be to appeal to nonconsequentialist considerations. It may be worth asking whether the injustice done by exploitative informing could satisfy this appeal. Third, we saw in Section 8.1 that an analogous result by Maher is less easy to challenge, at least insofar as we restrict ourselves to epistemic rather than practical challenges. However, if we understand exploitative informing as perpetrating a distinctively epistemic type of injustice, those who think that epistemic injustice can compete with other traditionally epistemic considerations (Basu 2019; Fricker 2007) may find in this thought ample epistemic grounds for the permissibility of resisting exploitative informing.[6] Developing and testing these arguments at length might be an interesting project for future research.

## 9.2 Duties to gather evidence

An ongoing project of epistemologists is to explain why agents sometimes have duties to engage in costly processes of evidence-gathering (Hall and Johnson 1998; Woodard and Flores forthcoming). Our discussion of exploitative informing suggests two new grounds for duties to gather evidence. The first ground is self-regarding: since exploitative informing rests on the asymmetric power of informants to gather evidence, receivers can break or even reverse the power imbalance between themselves and exploitative informers by gathering and sharing evidence of their own.

A second ground for duties to gather evidence is other-regarding: agents can break the power that exploitative informers have over others by gathering evidence and sharing that evidence with others. For example, agents might challenge the control that state-run media exercises over political narratives in authoritarian regimes by founding independent

[6] Others (Dotson 2019; Thorstad 2022) may be less concerned about the epistemic lens.

news outlets (Gehlbach and Sonin 2014; Stanley 2015). This second suggestion interacts with a broader consensus that competition in information markets tends to increase the quality of information provision (Bergemann and Bonnatti 2019; Mligrom and Roberts 1986; Gentzkow and Kamenica 2017). We might attempt to ground other-regarding duties to gather evidence in a broader view on which gathering and sharing evidence not only prevents others from being exploited, but also significantly improves the quality of information they will receive from many or all sources.

### 9.3 Generalizations

The discussion in this paper took a tightly constrained view of exploitative informing. All of the cases discussed had most or all of the five exploitative features discussed in Sections 3-7. In most cases, they had these features to a very high degree: for example, the acts we considered were maximally surplus hogging, providing no value at all to the receiver. It might be productive to consider cases where some of these exploitative features are lacking or are instantiated to a lesser degree. Are these cases still rightly regarded as exploitative? Does it remain permissible to resist information provision in these cases? And would we still want to say that information provision in these cases constitutes a type of epistemic injustice? Working through a broader range of cases should aid in understanding the nature, scope and norms of exploitative informing.

## References

Adams, Ernest and Rosenkrantz, Roger. 1980. "Applying the Jeffrey decision model to rational betting and information acquisition." *Theory and Decision* 12:1–20.

Basu, Rima. 2019. "Radical moral encroachment: The moral stakes of racist beliefs." *Philosophical Issues* 29:9–23.

Bergemann, Dirk and Bonnatti, Alessandro. 2019. "Markets for information: An introduction." *Annual Review of Economics* 11:85–107.

Bradley, Seamus and Steele, Katie. 2016. "Can free evidence be bad? Value of information for the imprecise probabilist." *Philosophy of Science* 83:1–28.

Brier, Glenn. 1950. "Verification of forecasts expressed in terms of probability." *Monthly Weather Review* 78:1–3.

Buchak, Lara. 2010. "Instrumental rationality, epistemic rationality, and evidence-gathering." *Philosophical Perspectives* 24:85–120.

Campbell-Moore, Catrin and Salow, Bernhard. 2020. "Avoiding risk and avoiding evidence." *Australasian Journal of Philosophy* 98:495–515.

Dorst, Kevin. 2019. "Lockeans maximize expected accuracy." *Mind* 128:175–211.

Dotson, Kristie. 2014. "Conceptualizing epistemic oppression." *Social Epistemology* 28:115–38.

—. 2019. "Tales from an apostate." *Philosophical Issues* 29:69–83.

Dreier, Jamie. 2011. "In defense of consequentializing." *Oxford Studies in Normative Ethics* 97–119.

Dworczak, Piotr and Martini, Giorgio. 2019. "The simple economics of optimal persuasion." *Journal of Political Economy* 127:1993–2048.

Ely, Jeffrey. 2017. "Beeps." *American Economic Review* 107:31–53.

Ely, Jeffrey and Szydlowski, Martin. 2020. "Moving the goalposts." *Journal of Political Economy* 128:468–506.

Engel, Christoph. 2011. "Dictator games: a meta study." *Experimental Economics* 14:583–610.

Fazelpour, Sina and Danks, David. 2021. "Algorithmic bias: Senses, sources, solutions." *Philosophy Compass* 16:e12760.

Foley, Richard. 2009. "Belief, degrees of belief, and the Lockean thesis." In Franz Huber and Christoph Schmidt-Petri (eds.), *Degrees of belief*, 37–47. Synthese Library.

Fricker, Miranda. 2007. *Epistemic injustice*. Oxford University Press.

—. 2017. "Evolving concepts of epistemic injustice." In Ian James Kidd, José Medina, and Gaile Pohlhaus Jr. (eds.), *Routledge handbook of epistemic injustice*, 53–60. Routledge.

Gehlbach, Scott and Sonin, Konstantin. 2014. "Government control of the media." *Journal of Public Economics* 118:163–71.

Gentzkow, Matthew and Kamenica, Emir. 2017. "Competition in persuasion." *Review of Economic Studies* 84:300–22.

Good, I.J. 1966. "On the principle of total evidence." *British Journal for the Philosophy of Science* 17:319–321.

Hall, Richard and Johnson, Charles. 1998. "The epistemic duty to seek more evidence." *American Philosophical Quarterly* 35:129–39.

Hoffman, Elizabeth, McCabe, Kevin, and Smith, VErnon. 1996. "Social distance and other-regarding behavior in dictator games." *American Economic Review* 86:653–60.

Howard, Ronald. 1966. "Information value theory." *IEEE Transactions on Systems Science and Cybernetics* 2:22–6.

Joyce, James. 1998. "A nonpragmatic vindication of probabilism." *Philosophy of Science* 65:575–603.

Kadane, Joseph, Schervish, Mark, and Seidenfeld, Teddy. 2008. "Is ignorance bliss?" *Journal of Philosophy* 105:5–36.

Kamenica, Emir. 2019. "Bayesian persuasion and information design." *Annual Review of Economics* 11:249–72.

Kamenica, Emir and Gentzkow, Matthew. 2011. "Bayesian persuasion." *American Economic Review* 101:2590–2615.

Maher, Patrick. 1990. "Why scientists gather evidence." *British Journal for the Philosophy of Science* 41:103–119.

Medina, José. 2017. "Varieties of hermeneutical injustice." In Ian James Kidd, José Medina, and Gaile Pohlhaus Jr. (eds.), *Routledge handbook of epistemic injustice*, 41–52. Routledge.

Mligrom, Paul and Roberts, John. 1986. "Relying on the information of interested parties." *RAND Journal of Economics* 17:18–32.

Muñoz, Daniel. 2021. "The rejection of consequentializing." *Journal of Philosophy* 118:79–96.

Neth, Sven. forthcoming. "Rational aversion to information." *Philosophy of Science* forthcoming.

Portmore, Douglas. 2011. *Commonsense consequentialism: Wherein morality meets rationality*. Oxford University Press.

Schroeder, Mark. 2007. "Teleology, agent-relative value, and 'good'." *Ethics* 117:265–95.

Seidel, Christian. 2019. "New wave consequentialism: An introduction." In Christian Seidel (ed.), *Consequentialism: New directions, new problems*, 1–30. Oxford University Press.

Shimony, Abner. 1988. "An adamite derivation of the calculus of probability." In James Fetzer (ed.), *Probability and causality: Essays in honor of Wesley C. Salmon*, 79–89. Synthese Library.

Stanley, Jason. 2015. *How propaganda works*. Princeton University Press.

Tenenbaum, Sergio. 2014. “The perils of honest consequentializing.” *Philosophy and Phenomenological Research* 88:233–40.

Thorstad, David. 2022. “There are no epistemic norms of inquiry.” *Synthese* 200:410.

Van Fraassen, Bas. 1983. “Calibration: A frequency justification for personal probability.” In Robert Cohen and Larry Lauden (eds.), *Physics, philosophy, and psychoanalysis: Essays in honor of Adolf Grunbaum*, 295–315. D. Reidel.

Verma, Sahil and Rubin, Julia. 2018. “Fairness definitions explained.” In *Proceedings of the International Workshop on Software Fairness*, 1–7. Association for Computing Machinery.

Woodard, Elise and Flores, Carolina. forthcoming. “Epistemic vigilance: In defense of epistemic norms on evidence-gathering.” *Philosophical Studies* forthcoming.