From “Deliberate Ignorance: Choosing Not to Know,” edited by Ralph Hertwig and Christoph Engel.
Strüngmann Forum Reports, vol. 29, Julia R. Lupp, series editor. Cambridge, MA: MIT Press. ISBN 9780262045599

10

The Zoo of Models of Deliberate Ignorance

Pete C. Trimmer, Richard McElreath, Sarah Auster, Gordon D. A. Brown, Jason Dana, Gerd Gigerenzer, Russell Golman, Christian Hilbe, Anne Kandler, Yaakov Kareev, Lael J. Schooler, and Nora Szech

Abstract

This chapter looks at deliberate ignorance from a modeling perspective. Standard economic models cannot produce deliberate ignorance in a meaningful way; if there were no cost for acquisition and processing, data could be looked at privately and processed perfectly. Here the focus is on cases where the standard assumptions are violated in some way. Cases are considered from an individual's perspective, without game-theoretic (strategic) aspects. Different classes of "not wanting to know" something are identified: aside from the boring case of the cost of information acquisition being too high, an individual may *prefer* to not know some information (e.g., when knowledge would reduce the enjoyment of other experiences) or may want to not *use* some information (e.g., relating to a lack of self-control). In addition, strategic cases of deliberate ignorance are reviewed, where obtaining information would also signal to others that information acquisition has occurred, and thus it may be better to remain ignorant. Finally, the possibility of deliberate ignorance emerging in population-level models is discussed, where there seems to be a relative dearth of models of the phenomenon at present. Throughout, the authors make use of examples to summarize different classes of models, ideas for how deliberate ignorance can make sense, and gaps in the literature for future modeling.

> Sometimes a man wants to be stupid if it lets him do a thing his cleverness forbids.
> — John Steinbeck, *East of Eden*

Group photos (top left to bottom right) Pete Trimmer, Richard McElreath, Anne Kandler, Jason Dana, Gerd Gigerenzer, Gordon Brown, Russell Golman, Christian Hilbe, Anne Kandler and Russell Golman, Richard McElreath, Sarah Auster, Yaakov Kareev, Gerd Gigerenzer, Nora Szech, Christian Hilbe, Pete Trimmer, Lael Schooler, Jason Dana, Nora Szech, Yaakov Kareev, Gordon Brown

From "Deliberate Ignorance: Choosing Not to Know," edited by Ralph Hertwig and Christoph Engel. Strüngmann Forum Reports, vol. 29, Julia R. Lupp, series editor. Cambridge, MA: MIT Press. ISBN 9780262045599

Introduction

The term *deliberate ignorance* implies that an agent has the option of obtaining some information and chooses not to. If there would be a great cost to acquiring the information, a deliberate choice to not acquire it would come as no surprise, so we restrict ourselves to considering cases where the information could have been acquired at a very small (or zero) cost to the agent, yet the agent chooses to remain ignorant.

In standard economic models of rational decision making, the expected "value of information" (i.e., knowing the information for free) can never be negative (Hirshleifer and Riley 1992). This is based on the assumption that the agent knows the underlying distribution of data with respect to the world so, although a particular sample may happen to be misleading, on average, the effect of receiving information will be non-negative. Simply put, standard models assume that an individual is in tune with their environment, so if information alters actions then, on average, it will improve the outcomes (and, if some data were deemed not to be useful, the individual could simply then continue as though they had not received it, which is different to ignoring the data in the first place). To understand this in a biological context, see McNamara and Dall (2010).

As the value of information is non-negative in standard models (and information is generally regarded as valuable for informing future decisions), the fact that individuals display deliberate ignorance (even when the cost of data acquisition is small) can seem surprising. It would be useful to understand when, and why, the phenomenon occurs. We will examine this by violating the standard assumptions of economic models in various ways, which is where theories and models come to bear.

It can be helpful to distinguish models from theories. Theories make assumptions about the world and from these assumptions derive predictions, which can be tested empirically. Models are abstract, simplified representations of the world or of theories about the world. Good theories will be predictively powerful (i.e., consistent with empirical data, but not generally consistent with *any* imaginable data), broadly applicable across a wide range of situations, and have assumptions that are parsimonious. Good models are realistic enough to describe the essential features of the world that the model is trying to capture, yet simple enough to give us insight. Models cannot perfectly describe the world in all its complexity; they are useful if they help us understand a particular aspect of the world by abstracting away from irrelevant details. Formalizing a theory as a mathematically precise theoretical model can help ensure that we fully understand the theory, that we are aware of its assumptions (e.g., converting implicit assumptions of a verbal theory to explicit), and that we can derive from them unambiguous predictions (it can be difficult to know the predictions of a complex theory without a model).

There are many types of models (e.g., descriptive, predictive, normative) and each is, by definition, an abstraction, so no model is a perfect representation

From "Deliberate Ignorance: Choosing Not to Know," edited by Ralph Hertwig and Christoph Engel. Strüngmann Forum Reports, vol. 29, Julia R. Lupp, series editor. Cambridge, MA: MIT Press. ISBN 9780262045599

of the real world. Consequently, the choice of model should depend on its intended purpose. Models have many potential benefits, including

- exposing the logic of a situation,
- making new, or more accurate, predictions,
- making predictions independent of the theorizer,
- helping to guide worthwhile empirical work (both for testing and improving on existing models and theories), and
- informing discourse on the likely effect of changes to the system (e.g., for policy planning).

Not all models are good; some models that seemingly explain a problem merely redescribe the data with regard to a new term which assumes the effect.[1] It is also easy for modeling to create "just-so" stories, providing seeming explanations for an effect, but with no other capability. To avoid just-so stories, the key is to find predictions from the model that we did not know beforehand. This allows the model to be tested. It is also worth noting that models often only produce predictions over a limited range of conditions. Although this may initially seem like a limitation, the fact that a model may prescribe different outcomes over different ranges means that the model has predictive power.

While it is generally not possible to show that a model is "right" (just because it produces correct predictions does not mean that it will do so in all cases, forevermore), it *is* possible to show that one is wrong, and it is often possible to contrast the predictions of models against one another. Models of deliberate ignorance could be used to

- show *why* deliberate ignorance exists,
- help with our ability to infer cases of deliberate ignorance (e.g., from behavior or physiological measurements), and
- understand the implications of policy changes (on whether particular information will be avoided, for instance) by modifying existing parameters or introducing new aspects to an existing model.

From a biological perspective, the question of "why" an organism displays deliberate ignorance can potentially be explained in four ways: mechanism (causation), function (adaptive value), ontology (development), and phylogeny (Tinbergen 1963). Each of these approaches can have their own models, and even if one model is perfect in a particular role, it may not necessarily help in another role. For instance, a mechanistic model of how the brain operates and results in an individual choosing to ignore information in a particular situation may be unlikely to help explain why that kind of brain evolved (from a functional perspective of fitness maximization in the species).

[1] Molière's parody to the question, "Why does opium make people sleepy?" is "because of its dormative properties." Similarly, the fact that people give money to others in the dictator game has been "explained" by other-regarding motives, which is very close to redescription.

From "Deliberate Ignorance: Choosing Not to Know," edited by Ralph Hertwig and Christoph Engel. Strüngmann Forum Reports, vol. 29, Julia R. Lupp, series editor. Cambridge, MA: MIT Press. ISBN 9780262045599

Perhaps the most intriguing of these cases is the functional question of why natural selection would have favored individuals with a decision-making system which ignored cheap but potentially useful data (or, more precisely, deliberate ignorance in cases where the expected value of the information minus the cost of acquisition is greater than the expected value of not having the information). When dealing only with the data itself (without any signal being available to others of knowing whether an individual has obtained the data), the functional reasons for deliberate ignorance existing are

- the cost of gathering the data is prohibitive,
- the cost of storing and processing the data are prohibitive, and
- processing of information (e.g., not being able to switch off automatic processes, resulting in suboptimal actions) is suboptimal.

When data acquisition also signals to others that data have been received, there is a fourth strategic reason for choosing to remain ignorant, which we discuss below (see section on "Interpersonal Strategic Perspectives").

We take an example-based approach to discussing various classes of models and distinguish two within-individual reasons for deliberate ignorance, irrespective of others. First, we discuss cases where an individual would *prefer* to not know some information (e.g., when knowing would reduce the enjoyment of other experiences). We then discuss cases where an individual would want to not *use* some information (e.g., relating to a lack of self-control). Thereafter, we turn to the strategic cases of deliberate ignorance: the effect of signaling that information acquisition has occurred. Finally, we explore the possibility of deliberate ignorance emerging in population-level models.

Preferring to Not Know Information

It is easy to list cases where an individual would prefer to not know some information. When reading a murder mystery, for instance, it would be easy to flick to the last page (and learn the culprit's identity) before reading the rest. For most people, doing so would reduce their enjoyment of the book; they would prefer to remain deliberately ignorant until reading to the end. Many hedonic reasons may be supplied for avoidance of such information. For some it may relate to the feeling of suspense, whereas others may enjoy trying to work something out. In some cases, such as when to hear the punch line of a joke, there is likely to be fairly universal agreement that deliberate ignorance is best. In other cases, this aspect can differ significantly between people. For instance, although some would like to know how a magic trick is done or to understand how a rainbow is formed, others may prefer (and thus deliberately choose) not to know. Such choices can depend on how an individual perceives the expected payoffs of knowledge and their subsequent interactions with the world. In this section, we discuss models where an individual seeks to maximize their

From "Deliberate Ignorance: Choosing Not to Know," edited by Ralph Hertwig and Christoph Engel. Strüngmann Forum Reports, vol. 29, Julia R. Lupp, series editor. Cambridge, MA: MIT Press. ISBN 9780262045599

expected level of "happiness." There are numerous models that can be used to explain deliberate ignorance under such circumstances; several are based on an individual's subjective utility.

Subjective Utility

The standard approach to decision making within economics is known as subjective expected utility theory (SEUT). According to SEUT, we as theorists can, if we make certain assumptions about people's rationality, make sense of a person's choices between possible outcomes if we assume that people behave as if they possess (a) stable utility functions and (b) beliefs about the probabilities of different outcomes.

More specifically, consider an event with just two possible outcomes, x_1 and x_2, and assume that an individual believes that the probabilities of these outcomes are $p(x_1)$ and $p(x_2)$, respectively. The subjective expected utility associated with the event will then simply be

$$u(x_1) \times p(x_1) + u(x_2) \times p(x_2), \tag{10.1}$$

where $u(x_i)$ is the utility of x_i for that person. The SEUT approach then assumes that the action (i.e., event) with the highest subjective expected utility is chosen. Note that SEUT assumes that individuals behave "as if" they are calculating the best option all the time, but the models often say nothing about the process by which such decisions are reached.

Perspectives on Belief-Based Utility

Bayesian updating is one of the standard approaches used to represent learning and optimal decision making in economic models; this approach fits with SEUT. Recent models from behavioral economics on deliberate ignorance can be divided in two groups, depending on whether they rely on Bayesian updating or not.

In the first case, agents are assumed to be Bayesian updaters. They may decide to avoid or ignore information and thus stick to their Bayesian prior to reduce problems of self-control, keep a halfway decent self-image, or stick to a not-too-drastic belief about their health status. Examples include Bénabou and Tirole (2002), Carrillo and Mariotti (2000), and Mariotti et al. (2018), as well as some models in which beliefs enter the utility function directly such as those of Caplin and Leahy (2001, 2004) or Schweizer and Szech (2018). In this class of models, agents will always respect the rules of Bayesian updating. As an illustration, take the example of Huntington disease: An agent knows that one parent carries the genetic mutation for the disease, while the other parent does not. The probability that the agent has the mutation is 50%. As she is Bayesian, this is also her belief of having it. If she takes a perfectly revelatory test, she will learn that there is the mutation in her blood (leading to disease) or not.

From "Deliberate Ignorance: Choosing Not to Know," edited by Ralph Hertwig and Christoph Engel. Strüngmann Forum Reports, vol. 29, Julia R. Lupp, series editor. Cambridge, MA: MIT Press. ISBN 9780262045599

When this agent thinks about getting information from the test versus ignoring it, she can only end up in situations where her belief about getting Huntington disease is 0%, 50%, or 100%. If she decides to ignore information, she sticks to the Bayesian prior of 50%.

This is very much in contrast to the second class of models, such as Brunnermeier and Parker (2005) or Gollier and Muermann (2010), in which agents "optimize" their beliefs. Consider again the example of Huntington disease. If the agent has not been tested yet, she can choose her beliefs to be anything from 0% to 100%. Thus, she may want to bias her beliefs optimistically and deviate substantially from the Bayesian prior of 50%. The latter models thus provide a lot of leeway to design (i.e., bias) beliefs, as Bayesian rules do not have to be followed.

An intermediate solution is proposed by Golman and Loewenstein (2018), who put forward an information gap belief-based utility model in which the impact of beliefs on utility depends on the attention paid to those beliefs. They assume that getting information attracts attention to the affected beliefs, with more surprising information attracting more attention. Golman et al. (2019) analyze the predictions of this model for information acquisition and avoidance. Information that may produce beliefs that are unpleasant to think about can have disutility because it forces people to pay more attention to beliefs that they do not want to attend to. This disutility is traded-off against the pleasure of satisfying curiosity and the instrumental value of the information. The model predicts that when beliefs are sufficiently unpleasant to think about, a person will prefer to remain deliberately ignorant, and as the intrinsic valence of that belief gets worse, the person would be willing to pay even more to remain ignorant. The model also makes predictions about when curiosity will overcome deliberate ignorance and cause a person to seek out information.

Contrasting Two Utility-Based Models

In a standard utility model, there are states of the world $\theta_i \in \Theta$ with probability $p(\theta_i)$ and choices between strategies (or actions) $s_j \in S$ that map from the set of states Θ into a set of material outcomes X, with a utility function defined on X. The standard value of information is

$$\sum_{\Theta} p(\theta_i) \max_{s_j \in S} u\left(x_{ij}\right) - \max_{s_j \in S} \sum_{\Theta} p(\theta_i) u\left(x_{ij}\right) \geq 0. \tag{10.2}$$

In belief-based utility models, beliefs about the state of the world enter the utility function, but the value of information can still be modeled as the expected utility of the posterior beliefs (including the utility of the choices made contingent on those beliefs) minus the utility of the prior belief (including the utility of the choice made given that prior):

From "Deliberate Ignorance: Choosing Not to Know," edited by Ralph Hertwig and Christoph Engel. Strüngmann Forum Reports, vol. 29, Julia R. Lupp, series editor. Cambridge, MA: MIT Press. ISBN 9780262045599

$$\sum_{\Theta} p(\theta_i) \max_{s_j \in S} u\left(\boldsymbol{p}^{\theta_i}, x_{ij}\right) - \max_{s_j \in S} u\left(\boldsymbol{p}, \boldsymbol{x}_j\right), \tag{10.3}$$

where $\boldsymbol{p}$ is the prior belief about states of the world, $\boldsymbol{p}^{\theta_i}$ is belief after learning that the state is θ_i (i.e., the degenerate distribution on this state), and $\boldsymbol{x}_j$ is the (prior) distribution over outcomes that would result from choosing strategy s_j. We may rewrite this as

$$\sum_{\Theta} p(\theta_i) U\left(\boldsymbol{p}^{\theta_i}, x_{ij} \mid S\right) - U\left(\boldsymbol{p}, \boldsymbol{x}_j \mid S\right), \tag{10.4}$$

where $U\left(* \mid S\right) = \max_{s_j \in S} u\left(*\right)$.

In the information gap model, attention enters the utility function, and if π denotes beliefs about answers to questions and about the distribution over outcomes (i.e., π takes the place of $\boldsymbol{p}$, $\boldsymbol{x}_j$), the value of information from answering a question becomes

$$\sum_{\mathrm{i}} \pi(A_i) U\left(\pi^{A_i}, \boldsymbol{w}^{A_i} \mid S\right) - U\left(\pi, \boldsymbol{w} \mid S\right), \tag{10.5}$$

where the A_i are the possible answers to the question, $\boldsymbol{w}$ is the attention placed on each question before getting any information, and $\boldsymbol{w}^{\theta_i}$ is the attention placed on each question after finding out that the answer is A_i (Golman et al. 2019). The information gap model assumes that the attention weight vector $\boldsymbol{w}$ depends on the importance of the various questions (which is modeled in terms of the spread of the utilities that would result from different answers) and on the salience of the various questions (which is not modeled at all); the attention weight $\boldsymbol{w}^{A_i}$ additionally depends on the surprise associated with finding out answer A_i (which is modeled in terms of Kullback-Leibler divergence). The model also assumes a specific form for the utility function:

$$u\left(\pi, \boldsymbol{w}\right) = \sum_{x \in X} \pi_X\left(x\right) v_X\left(x\right) + \sum_k w_k \left(\sum_i \pi_k\left(A_{ki}\right) v_k\left(A_{ki}\right) - H\left(\pi_k\right) \right), \tag{10.6}$$

where H is the entropy function (a measure of how uncertain each belief is) and k indexes the various questions that the person is aware of (Golman and Loewenstein 2018).

Consider, for example, an opportunity to get tested for HIV with the assumptions that this is the only question that the individual is aware of, that the individual would choose to take medicine if the test is positive, and that they would choose to not take medicine if the test is negative or if he remains deliberately ignorant (see Golman and Loewenstein 2018). The value of information becomes

$$\pi\left(\mathrm{Pos}\right) u\left(\pi^{+,\mathrm{medicine}}, w^{+}\right) + \pi\left(\mathrm{Neg}\right) u\left(\pi^{-}, w^{-}\right) - u\left(\pi, w\right). \tag{10.7}$$

Without loss of generality, consider the material value of not having HIV to be 0, and also assume that not having HIV has neutral belief valence 0 (i.e.,

From "Deliberate Ignorance: Choosing Not to Know," edited by Ralph Hertwig and Christoph Engel. Strüngmann Forum Reports, vol. 29, Julia R. Lupp, series editor. Cambridge, MA: MIT Press. ISBN 9780262045599

the person does not mind or enjoy thinking about not having HIV). Then $u\left(\pi^{-}, w^{-}\right) = 0$.

Letting the material value of having HIV and taking medicine be v_{x_M} and the valence of believing that one has HIV be v_H, we get

$$u\left(\pi^{+,\text{medicine}}, w^{+}\right) = v_{x_M} + w^{+} v_H . \tag{10.8}$$

Lastly, letting the material value of having untreated HIV be v_{x_H} and letting $p = \pi\left(\text{Pos}\right)$, we get

$$u\left(\pi, \boldsymbol{w}\right) = p v_{x_H} + w\left(p v_H - H\left(p\right)\right). \tag{10.9}$$

Putting this together, a person would choose to be deliberately ignorant of their HIV status if

$$p\left(v_{x_M} + w^{+} v_H\right) < p v_{x_H} + w\left(p v_H - H\left(p\right)\right). \tag{10.10}$$

Rearranging terms, the condition for deliberate ignorance is

$$p(v_{x_M} - v_{x_H}) + wH\left(p\right) < (w^{+} - w)\left(-p v_H\right). \tag{10.11}$$

Interpreting Eq. 10.11, the instrumental value of the information $p(v_{x_M} - v_{x_H})$ plus the intrinsic value of reducing uncertainty (or satisfying curiosity) $wH(p)$ needs to be less than the benefit of not increasing attention on a negative-valence belief $(w^{+} - w)(-pv_H)$. The assumption that surprise attracts attention implies that $(w^{+} - w) > 0$. The prediction of whether or not the person chooses to be deliberately ignorant depends on how much difference it makes to take medicine when HIV positive $(v_{x_M} - v_{x_H})$, on how unpleasant it is to think about being HIV positive (v_H), and on how much attention the person was initially paying (w), which itself depends on how salient the question was and on its importance. The model predicts that if thinking about being HIV positive is sufficiently bad, the person will choose to be deliberately ignorant, and that given fixed values of how unpleasant it is to think about being HIV positive and on how much taking medicine helps in that case, a person could choose to be deliberately ignorant if the question is not initially salient (and thus attracts little attention), but could change his mind and choose to become informed if the question becomes highly salient.

In a model based on optimism, deliberate ignorance may arise from a person choosing to hold optimistic beliefs in the absence of information but being unable to maintain optimistic beliefs after getting information. Following Oster et al.'s (2013) analysis of Huntington disease testing, we can use Brunnermeier and Parker's (2005) model of optimism to analyze HIV testing. Accordingly, if a person chooses to be deliberately ignorant, he can choose his belief (i.e., the probability q that he believes he has HIV) to reduce the anxiety of having it at the cost of then potentially mistreating it, even though he makes the decision of whether to remain deliberately ignorant in the first place in some sense knowing the true probability p that he actually has HIV. While in principle he could

From "Deliberate Ignorance: Choosing Not to Know," edited by Ralph Hertwig and Christoph Engel. Strüngmann Forum Reports, vol. 29, Julia R. Lupp, series editor. Cambridge, MA: MIT Press. ISBN 9780262045599

choose any value of q, his choice really comes down to whether to choose $q = q^*$, the minimum level of risk that would induce him to seek treatment, or $q = 0$ (no risk). A person who chooses $q = 0$ has no anxiety but has a probability p of having untreated HIV, which has value v_{x_H}, so the expected utility of choosing $q = 0$ is pv_{x_H}. A person who chooses $q = q^*$ will get the treatment and thus has a p chance of having HIV with treatment, which has value v_{x_M} and a $1 - p$ chance of getting unnecessary treatment despite not having HIV, which has value v_{x_N}. This belief choice also leads to anxiety from anticipating these outcomes with probabilities q^* and $1 - q^*$ respectively. Thus, the expected utility of choosing $q = q^*$ is

$$pv_{x_M} + (1-p)v_{x_N} + \delta\left(q^* v_{x_M} + (1-q^*)v_{x_N}\right), \tag{10.12}$$

where δ is the weight placed on the belief-based utility (i.e., anxiety). (Note that $q^* v_{x_M} + (1-q^*)v_{x_N} = q^* v_{x_H}$ because the person must be indifferent about treatment at q^*.)

The person chooses $q = 0$ if

$$pv_{x_H} > pv_{x_M} + (1-p)v_{x_N} + \delta \frac{v_{x_H} v_{x_N}}{v_{x_H} + v_{x_N} - v_{x_M}}. \tag{10.13}$$

If the person gets the HIV test, he can no longer choose his belief. Instead his expected utility, conditional on proper treatment, is $pv_{x_M}(1+\delta)$.

The person will choose to be deliberately ignorant of the test results if

$$pv_{x_M}(1+\delta) < \max\left\{ pv_{x_H},\ pv_{x_M} + (1-p)v_{x_N} + \delta \frac{v_{x_H} v_{x_N}}{v_{x_H} + v_{x_N} - v_{x_M}} \right\}. \tag{10.14}$$

The model predicts that if the thought of having HIV (even if treated) is sufficiently negative, then placing enough weight on anticipatory utility (i.e., being sufficiently anxious) will cause somebody to avoid the test result (i.e., choose to be deliberately ignorant).

Finally, we note that although the subjective utility approach assumes that utility functions may be inferred, there has been no attempt here to tie this in with evolving those functions (from a functional perspective of the utility curves being adaptively beneficial). Natural selection acts on our behaviors, irrespective of how we feel about things (individuals who constantly feel sad have the same expected fitness as those who constantly feel happy, if the actions of each are the same). So, while we care about mental happiness, pain and so on, these are only adaptive insomuch as they assist our mental drives in guiding us toward adaptive behavior in certain situations.

Links with Forgetting and Heuristics

Schooler (this volume) outlines deliberate strategies that people might use to prevent the retrieval of emotionally disturbing information. This is akin to the

From "Deliberate Ignorance: Choosing Not to Know," edited by Ralph Hertwig and Christoph Engel. Strüngmann Forum Reports, vol. 29, Julia R. Lupp, series editor. Cambridge, MA: MIT Press. ISBN 9780262045599

"emotion-regulation and regret-avoidance" function that Hertwig and Engel (this volume) ascribe to deliberate ignorance (see also Gigerenzer and Garcia-Retamero 2017). One strategy is to encode new memories that interfere with the retrieval of the disturbing memories. A second approach is to exploit human memory's propensity to confuse imagined events with real ones (Loftus 1997). In essence, rather than having accurate memories of the past, it is better for our emotional well-being to mask the bad experiences with false memories (thus becoming deliberately ignorant of the past). Techniques to reconsolidate negative memories with more positive ones are being developed to treat people with posttraumatic stress disorder (Gardner and Griffiths 2014). However, to the best of our knowledge, precise computational models for the construction of false memories, whether beneficial or not, are rare (an exception is Hoffrage et al. 2000). This approach also suffers from the same difficulty as mentioned above, of not linking with the functional aspect of whether such processes make sense from an adaptive perspective.

Making Use of Ignorance

We conclude this section on a somewhat different tack, by noting that the condition of ignorance can itself be informative, as ignorance can correlate with what one wants to know. For instance, the recognition heuristic (Goldstein and Gigerenzer 2002) has been shown to be highly effective in some scenarios. Consider the prediction of the outcomes of tennis matches at a major event. If a spectator has heard of one player but not their opponent, the recognition heuristic predicts that the player whose name is recognized will win. Mere recognition has been shown to predict as well as, or better than, the ATP rankings and Wimbledon experts (Serwe and Frings 2006). (Note, however, that ATP rankings are not set up purely to predict match outcomes; the rankings include a recency bias to encourage players to play more matches.) The recognition heuristic has also been successfully used to ignore information deliberately while investing in portfolios of stocks that reflect the limited name recognition of firms (Borges et al. 1999; Ortmann et al. 2008). The heuristic works well in situations where a lack of recognition has high predictive power. Thus, heuristics can make use of a lack of knowledge; ignorance (lack of recognition) in the recognition heuristic is, itself, information. When that cue of ignorance has greater validity than other potential cues, it is theoretically possible for an individual to benefit by preferring to remain ignorant rather than recognize additional cases (e.g., when they recognize half the players in a tennis tournament).

This outcome does not contradict the findings of the standard economic model as it violates the standard assumptions by assuming a form of bounded rationality; recognition is binary, in contrast with standard models which would assume graded information levels (e.g., of how often a player had been seen before) and full use of all available information.

From "Deliberate Ignorance: Choosing Not to Know," edited by Ralph Hertwig and Christoph Engel. Strüngmann Forum Reports, vol. 29, Julia R. Lupp, series editor. Cambridge, MA: MIT Press. ISBN 9780262045599

Ignorance (or partial ignorance) can also be beneficial when it comes to choosing what to attempt. For instance, someone trying to climb the academic ladder might benefit from not recognizing that all of the current professors are of the opposite sex. Such information could be disheartening and cause the person to believe that a professorship is most unlikely, a result that would then be self-fulfilling. It is conceivable that someone might be subconsciously aware of such a fact and then "choose" (i.e., consciously, and thus deliberately) not to look into data on the topic. If someone's work ethic could be influenced by such data, this strategy may be beneficial.

Wanting to Not Use Information

We now turn to cases where making use of information is expected to lead to worse outcomes, as in situations where processing data is automatic. Under these circumstances, it obviously makes sense to be deliberately ignorant. Some cases of deliberate ignorance are imposed at a group level (e.g., what a jury is allowed to know about a defendant); here we focus on choices at an individual level.

Overfitting and Forgetting

It has long been recognized that it is easy to "overfit" data. Given a set of instances from which to learn, a model can be fitted using features, by finding the set of parameters which best predicts the outcome data from the feature data. However, when used in an unsophisticated way, this approach typically tries to make too much use of the feature data and, to make predictions, it would be better to deliberately ignore some of the features because the "true" validity of the information is unknown. Consider, for example, a wild salmon fishery where the goal is to know the structure of the true biological model of a salmon fishery's population. Such a model will have many unknown parameters, because salmon are complicated animals with complicated life histories. If we are unable to gather enough data to estimate accurately all of the different parameters, we might make better decisions by using a simpler model that ignores information. If, however, we knew the values of the parameters with sufficient accuracy, then ignoring this information would not lead to better decisions. Thus, the value of deliberate ignorance arises from ignorance of something else.

Kareev (2012) shows that a correlation is more likely to be detected as sample size (or memory) decreases. In contrast to the paragraph above, this does not arise from not knowing something else about the system. When decisions are discrete, there may be nothing that we can tell the agent that would mean that acquiring a few more samples would improve decisions. However, like the salmon example, it is also a consequence of how bias and variance

From "Deliberate Ignorance: Choosing Not to Know," edited by Ralph Hertwig and Christoph Engel. Strüngmann Forum Reports, vol. 29, Julia R. Lupp, series editor. Cambridge, MA: MIT Press. ISBN 9780262045599

jointly influence patterns of error. There are numerous techniques for eliminating this problem, such as using simple heuristics, take-one-out (or, more generally, n-fold) validation or, in the case of decision trees, various types of pruning (Mitchell 1997). When data is handled appropriately, there is no harm in receiving such data (if it is free to acquire and process) as it will not be overused, so there is no benefit in being deliberately ignorant. However, there are a couple of obvious caveats:

- If there is a cost to acquiring or processing the data, then it can be better to choose to avoid it.
- If processing of the data is automatic (i.e., it cannot be ignored) and is not always suitable, then it can be better to avoid it.

It is the second of these cases that we focus on here, as the more interesting. One of the reasons that using data might be inappropriate is that it may be out of date, thus it is no longer beneficial. Schooler (this volume) explores how memory processes, including forgetting, can achieve functions that Hertwig and Engel (this volume, 2016) have attributed to deliberate ignorance. There are well-developed computational models of human memory motivated by the observation that forgetting helps to prioritize important information and set aside information that is likely to be distracting (Anderson and Milson 1989). Beyond removing potentially interfering information, forgetting may be adaptive for specific purposes. For instance, Schooler and Hertwig (2005) implemented a model of the recognition heuristic and varied the forgetting parameter of their model. They showed that the recognition heuristic performed best at intermediate levels of forgetting. At low levels of forgetting, the model would likely recognize both options, whereupon the recognition heuristic could not be used. Similarly, at high levels of forgetting, neither name is likely to be recognized and again the recognition heuristic does not apply. However, there is no claim that forgetting is "deliberate" in any conscious sense in the basic model. However, when the world is changing, there is good reason to consciously discount old data: deliberately ignoring it, much like forgetting it, would then make plenty of sense as a way of making sure that it is not used.

Collider Bias

A valid reason to omit a variable from consideration is because including information can confound inference, regardless of how much data we have. The clearest example is known as collider bias. A *collider* is a variable that is a function of two (or more) other variables. Consider, for example, a lamp. Whether the lamp is "on" is a causal function of both the switch that controls it and the flow of electricity:

$$\text{switch} \rightarrow \text{lamp} \leftarrow \text{electricity}$$

From "Deliberate Ignorance: Choosing Not to Know," edited by Ralph Hertwig and Christoph Engel. Strüngmann Forum Reports, vol. 29, Julia R. Lupp, series editor. Cambridge, MA: MIT Press. ISBN 9780262045599

The lamp is a collider of the switch and electricity. Once we know the state of the lamp, it provides information about the switch and the electricity. If the lamp is on, we know that both the switch and the electricity are on as well. If the lamp is off, then either the switch or electricity (or both) are off. If we know the lamp is off and the switch is on, then the electricity must be off. The point is that while causation flows in one direction, from causes (switch or electricity) to results (lamp), statistical information can flow in all directions.

What does this have to do with confounding? Suppose we wish to learn about the relationship between education (E) and wages (W). How much does education influence (cause) wages, $E \rightarrow W$? Suppose also that education and wages jointly influence hobbies (H), like sky diving or watching football. Now H is a collider of E and W. As a result, if we learn someone's hobbies and their education, we also learn something about wages, in the same way that knowing whether the lamp is on and the switch is on tells us about the electricity.

If we then regress W on E, including H as a covariate, it will result in a confounded estimate of the causal influence of E on W. Why? Because as soon as we condition on H (learn about H), statistical association flows along the path $E \rightarrow H \leftarrow W$ and biases our inference. We end up polluting the path $E \rightarrow W$ with information from the other path. If, instead, we omit H from consideration, no information flows along the path $E \rightarrow H \leftarrow W$ because, if we do not know H, then E tells us nothing about W. There may be a number of variables like H (e.g., marriage status or number of children), and including any one of them as a "control" variable would bias inference, regardless of how much data we collect. Therefore, the reason for ignoring a control variable is distinct from overfitting concerns.

An interesting feature of the collider bias example is that it requires enough knowledge of the causal model in order to stimulate deliberate ignorance of the collider (hobbies in the example). Therefore, a person practicing deliberate ignorance of, for instance, hobbies already knows (or thinks they know) more about hobbies than a person who might use hobbies in the analysis. So, the person is hardly ignorant about hobbies, in the abstract. Rather, they deliberately avoid gathering more information about them (as any such data should not be used).

Another feature is that it does not matter, in terms of inference, whether the collider is never learned or simply not used in the analysis. This is a property of many individual motives for deliberate ignorance. The recognition heuristic is a plausible exception: it is not easy to un-recognize something and recall that it was previously un-recognized. There, ignorance and nonuse are connected. When we turn later to interpersonal, strategic reasons for deliberate ignorance, the difference between ignorance and nonuse will be crucial.

From "Deliberate Ignorance: Choosing Not to Know," edited by Ralph Hertwig and Christoph Engel. Strüngmann Forum Reports, vol. 29, Julia R. Lupp, series editor. Cambridge, MA: MIT Press. ISBN 9780262045599

An Evolutionary Case, Using Collider Bias

Suppose that a species has evolved in a situation where they sometimes encounter potential cues, *X*s and *Z*s, and must choose their behavior based on their expectations of another parameter, *Y*, which can only be inferred at the time of decision. Let us first assume that there is collider bias, as shown in Figure 10.1, so the individual should only go by *X* without making use of *Z*.

Many such circumstances could be imagined. For instance, *X* could be any factor at a location that influences whether a nest should be built at that location, *Y* then represents whether it is good or bad to build a nest there (which cannot be directly perceived), and *Z* a factor influenced by the other two, such as the number of existing nests in the area. With zero costs for data, it would be fully rational to take in all information (*X*s and *Z*s) and then choose which data to make use of (as standard models of rationality typically assume). However, the costs associated with gathering information as well as the mental processes involved (e.g., building and maintaining memory banks, energy costs of processing information) mean that it can be best for an agent to have bounded rationality (in the sense of constraints on their mental capabilities). Consequently, rather than obtaining, storing, and mentally processing each case of *Z* that is encountered, natural selection may instead select for organisms which deliberately choose to ignore *Z*. (Note that there may also be a cost associated with ignoring an item; here we assume that this cost is smaller than the combination of the other costs if *Z* were to not be ignored, as seems likely in many biological cases). This then sets the stage for more extreme forms of deliberate ignorance to occur through environmental change.

Deliberate Ignorance Arising Through Environmental Change

In the modern world, deliberate ignorance may be displayed because recent environmental change has been occurring faster than our brains have been able to evolve.

Suppose that, in the ancestral environment, there was the possibility of obtaining a particular type of information, but at significant cost. For instance, by approaching spiders or snakes, and sometimes being bitten, one would learn

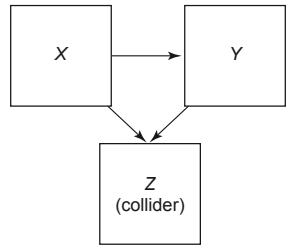

Figure 10.1 Causal diagram of the simplest case of collider bias.

From "Deliberate Ignorance: Choosing Not to Know," edited by Ralph Hertwig and Christoph Engel. Strüngmann Forum Reports, vol. 29, Julia R. Lupp, series editor. Cambridge, MA: MIT Press. ISBN 9780262045599

which were safe and which were not. Such information would be very useful, but generally it would be prohibitively costly to obtain. Instead, individuals may do much better to immediately avoid *all* snakes and spiders (regardless of their type) and largely only learn (what they could) about such animals from others. It makes sense, then, that psychological mechanisms should have evolved to steer clear of such dangers.

This, of course, is not an interesting case of deliberate ignorance, as the costs of obtaining such information are large. In the modern world, however, with glass cages and the like, snakes and spiders can be approached and learned about directly, without paying that cost of potentially being bitten. Yet many people still display a strong aversion to such creatures, even when they are clearly behind glass. Thus, although the cost of obtaining information about the animals would be very cheap to obtain (simply walking up to the glass and looking at them), many people[2] display deliberate ignorance in a fairly extreme way, by exercising immediate avoidance.

The ancestral reason for data avoidance can be any case where the costs of data acquisition and processing are higher than the expected benefits of having that information (e.g., any case of collider bias where there was a cost for the information). Having evolved the tendency to avoid that information, an environmental change that modifies expected payoffs can then result in individuals avoiding information where, if they were able to process it properly, otherwise it would be better to acquire the information.

Similarly, it is possible that *cues* relating to the expected cost of acquiring or processing the information (rather than the actual costs) may have altered from ancestral settings. Thus, it is potentially very easy for deliberate ignorance to arise through environmental change.

Blinding and Bidding

We have seen that factors such as overfitting and collider bias mean that it is easy to make use of data when one would do better to avoid it. As people will often tend to make use of a variable when they know it (e.g., names in peer review), the option of "blinding" can make sense. This can occur at different levels: between organizations, between sections of an organization, or between individuals.

Sealed Bids and Blind Trusts

In a sealed-bid auction, a number of bidders submit their bid simultaneously, without knowing each other's bid. In the simple case of bidding to buy an item

[2] This effect is not only observed in humans; numerous YouTube videos show cats reacting, in apparent terror, when they discover a cucumber that had been silently placed behind them. It seems highly unlikely that cucumbers ever posed a threat to cats in ancestral times, but it would have been highly adaptive for cats to leap away if a long green thing suddenly, and silently, appeared behind them.

From "Deliberate Ignorance: Choosing Not to Know," edited by Ralph Hertwig and Christoph Engel. Strüngmann Forum Reports, vol. 29, Julia R. Lupp, series editor. Cambridge, MA: MIT Press. ISBN 9780262045599

(e.g., an art piece), the envelopes are opened and the highest bidder wins the item (typically then being required to pay his/her bidding price or, in some auctions, the second highest price). Here, deliberate ignorance (in the form of a blind review) is not an issue.

Another common case, which *is* relevant to deliberate ignorance, is where bidding takes place for some large project (e.g., paving a road, building an office complex, contracting to develop some weapons system), and the call for bids is usually made by a public body (government, state, city). Generally, competitors have to submit a detailed proposal on what they propose to do, how they plan to do it, how much they would demand for the project, and so on. Once bids are submitted, they are opened, rated, and compared. Rating is usually based on a number of criteria (e.g., quality of the plan, price asked, previous experience, financial resources) whose relative weight is announced in advance. The final score is the weighted average of the scores on the various criteria. In such bids, one often does not want the identity of the bidder to interfere with the evaluations (e.g., you don't want political supporters of the mayor to get an unfair advantage). In such cases, deliberate ignorance—hiding the identity of the bidder—may help.

However, it may still benefit an individual to (surreptitiously) know the identity of each bidder. Thus, deliberate ignorance here is at a different level to that of the individual; it is at the organization level, shielding its assessment from knowledge in other parts of the organization.

We contrast this with a blind trust, where politicians, for example, shield themselves from knowledge of their financial investments. This is arguably done as a signal to others (especially voters) that the individual can be trusted to be acting on behalf of everyone, rather than taking self-interested actions. The action also serves the individual in the longer term by protecting them from criticism about their actions if others (correctly or incorrectly) accuse them of making self-interested decisions.

Collective Deliberate Ignorance

Prostate-specific antigen (PSA) screening (i.e., a blood test for the early detection of prostate cancer for men without symptoms) is not recommended by the U.S. Preventive Services Task Force (USPSTF) and other national health organizations. It has also been outright rejected by the Swiss Medical Board and other organizations, as well as Richard Ablin, the discoverer of PSA. Most health insurers do not pay for the test. The reason is that randomized studies have been unable to show that screening reduces the total mortality after ten years (i.e., no life is saved), yet many men are harmed (e.g., incontinence and impotence) through surgery or radiation treatments that follow a positive test. Nevertheless, many urologists still recommend PSA screening.

Studies have shown that most urologists do not know the benefits and harms of PSA screening and seem to prefer to remain ignorant, even though

From "Deliberate Ignorance: Choosing Not to Know," edited by Ralph Hertwig and Christoph Engel. Strüngmann Forum Reports, vol. 29, Julia R. Lupp, series editor. Cambridge, MA: MIT Press. ISBN 9780262045599

information is easily available on the USPSTF's website as well as through other national organizations (Gigerenzer 2014). Remaining ignorant can protect these physicians from being sued, as illustrated by the case of Daniel Merenstein, a U.S. physician who studied the evidence and informed a well-educated man about the pros and cons of PSA screening, after which the man declined the test. Unfortunately, a few years later the man got advanced prostate cancer and sued Merenstein for having informed him instead of performing the test. The man was awarded the maximum amount, despite the defense having brought in national experts who testified that the benefits of the PSA test are unproven but the harms are (Merenstein 2004).

Aside from the risk of being sued, there appear to be two additional reasons for deliberate ignorance related to PSA screening. First, in Germany, PSA tests and their downstream consequences (e.g., biopsies) result in about 25% of the average urologist's earnings. If urologists were to look up the scientific evidence, this might cause an internal conflict (or cognitive dissonance) between making money and their self-image of being a good doctor (Festinger 1957). This internal reason for deliberate ignorance is similar to the notion of maintaining identity (internal consistency). Second, a urologist who looks up the scientific evidence and presents it to other urologists in talks or writings may expect to be regarded as a troublemaker and choose deliberate ignorance over being disrespected.

Can Reading Block Original Thinking?

Another thought-provoking question concerns the following: Should academics should choose *not* to read the existing literature when starting to research a new topic? In some cases, reading the literature might bias one toward using existing paradigms and block original thinking (just as it can be harder to think of a particular word when a very similar word is already in mind). The extent to which it is worth reading the literature may depend on factors such as the quality of the researcher (arguably very high-quality individuals may do best to start afresh) and the number of other researchers who have already tried to make progress on the topic. One approach to modeling this would be to break a task into stages and assume that existing work has already addressed each one of the stages. A new researcher would then be able to choose whether to read the existing literature (in which case they will have some assumed heuristic for which stages to attempt to progress, based on the perceived progress with each stage) or have a clear run at each stage.

Of course, there are easy ways to build models in which individuals choose to ignore data, such as putting a time cost on reading the existing literature. However, by imposing such costs, these constraints shift the explanation to a boring category of the (obvious) benefits being outweighed by the (obvious) costs. This approach would instead be assuming that there was no cost for reading the existing literature, except that automatic heuristics would then make

From "Deliberate Ignorance: Choosing Not to Know," edited by Ralph Hertwig and Christoph Engel. Strüngmann Forum Reports, vol. 29, Julia R. Lupp, series editor. Cambridge, MA: MIT Press. ISBN 9780262045599

use of certain aspects of that existing work, which would then bias attempts to solve a task. The worth of such a model may therefore rest on whether the heuristics were worthwhile abstractions of reality.

Interpersonal Strategic Perspectives

Thus far we have considered deliberate ignorance in the context of an individual having the option of seeing data or not, without that choice being known to others. As we discuss here, when the choice of accessing data (or not) is known to others, there can be very good reasons for deliberate ignorance.

Consequences of Commitment

Hilbe and Schmid (this volume) provide an example of a two-player "envelope game" where it is better to be ignorant than knowledgeable, so long as the state of knowledge is known to the other player. The game provides a nice example of how parameters can govern the outcome of the system. In some settings, it is better to be knowledgeable, in which case deliberate ignorance would not be expected. In others, deliberate ignorance should exist if players are not able to communicate findings to one another. If they are able to, however, "cheap talk" would benefit both players. In a particular range of scenarios, deliberate ignorance emerges as the best strategy even when the players can communicate with one another. Models that predict qualitatively different phenomena can be usefully tested. However, biological examples of such a system are not easy to envisage.

Deliberate Ignorance as a Signal of Condition

If individuals differ in the extent to which they rely on information, deliberate ignorance can serve as a costly signal (Spence 1973). In that case, an individual's public decision not to learn information may help individuals to distinguish themselves from others.

To illustrate this point, consider an extension of the envelope game by Hilbe and Schmid (this volume), where there are two types of player 1: "favorable" players are generally more likely to have a low cooperation cost, whereas "unfavorable" players are more often in a high-cost environment. For some game parameters, the game has an equilibrium in which unfavorable players decide to learn the state of nature, whereas favorable players ignore it (Hilbe and Schmid, this volume). In this game, favorable players can thus use their ignorance as a signal that they can be trusted to cooperate. For this mechanism of deliberate ignorance to work, however, it is important that player 1's ignorance can be verified. If players could privately learn the state of nature, their deliberate ignorance would no longer serve as a costly signal.

From "Deliberate Ignorance: Choosing Not to Know," edited by Ralph Hertwig and Christoph Engel. Strüngmann Forum Reports, vol. 29, Julia R. Lupp, series editor. Cambridge, MA: MIT Press. ISBN 9780262045599

The game is related to the handicap principle (Zahavi 1975), which is often used to explain why individuals choose to take seemingly unnecessary risks. The driving factor in each case is that different types of individuals pay different costs. Zahavi's handicap principle shows that high-quality individuals can signal their quality by taking risks, thus influencing mate choice of others (to their benefit). In this envelope game, by signaling that they are going to remain deliberately ignorant, individuals can influence the choices of others, similarly to their benefit. In the handicap case, the high-quality individual takes the (somewhat unexpected) action of taking a risk; in the envelope game case, the "favorable" player 1 is advantaged by choosing the (somewhat unexpected) route of deliberate ignorance.

A similar signaling motive may explain other effects, such as why employers may choose not to monitor the work effort of their employees. From one perspective, monitoring should increase employee effort, because employees wish to avoid sanctions for shirking. From another perspective, employees may frame their relationship with management as a reciprocal relationship; in this case, monitoring may be interpreted as distrust and result in a reduction of effort. Note, however, that an employer might choose to fake their signal of whether they trust the employees, so the situation can be very complex. There is evidence that monitoring can reduce worker effort but the effect does not always arise (Dickinson and Villeval 2008). When it does, deliberate ignorance of worker effort may result in increased worker effort. Kareev and Avrahami (2007) also show that deliberate ignorance on the part of an employer may help to motivate less able workers to compete for bonuses.

Choosing Whether to Know Payoffs to Others

Table 10.1 shows the short-term payoffs to oneself and another individual when choosing between two options, which we simplistically label "up" and "down" (based on Dana et al. 2007, see also Dana, this volume). Faced with knowledge of the situation, many players may choose "up," sacrificing one unit of payoff in the immediate term to show that they are willing to help others. In contrast, Table 10.2 shows payoffs when there is no conflict: choosing "down" will be best for both players.

Suppose that an individual confronts one of the above situations and has the option of learning which situation (Table 10.1 or Table 10.2) they are facing. If they are ignorant of the situation being faced, they can choose "down" without

Table 10.1 Short-term payoffs in a situation with strong contrasts in payoffs to others (conflicting choice).

	Payoff to self	Payoff to other
Choosing "up"	5	5
Choosing "down"	6	1

From "Deliberate Ignorance: Choosing Not to Know," edited by Ralph Hertwig and Christoph Engel. Strüngmann Forum Reports, vol. 29, Julia R. Lupp, series editor. Cambridge, MA: MIT Press. ISBN 9780262045599

Table 10.2 Short-term payoffs in a situation with no conflicting choice.

	Payoff to self	Payoff to other
Choosing "up"	5	1
Choosing "down"	6	5

hesitation and thus ensure their largest possible payoff. But knowing which situation they face may cause internal conflict. Arguably, then, individuals may choose to be deliberately ignorant in such a situation. This is an example of an individual exercising their "moral wiggle room."

Models of Moral Wiggle Room

Research on the topic of moral wiggle room suggests that people sometimes choose to remain ignorant of the consequences of their desired actions, specifically because they would feel obliged to behave better (i.e., more altruistically or in accordance with social norms) if they knew the consequences with certainty (d'Adda et al. 2018; Dana et al. 2007; Freddi 2017; Grossman and Van Der Weele 2017; Serra-Garcia and Szech 2018). Dana et al. (2007) found that people avoid information about the consequences of their choices for other people so that they can make the choice that is in their own monetary self-interest. Of course, they could make the self-interested choice even if they were to discover that it would harm others, but then they would feel guilt. Consistent with a desire to avoid the information, Dana et al. (2007) found that only 56% of dictators chose to reveal the recipient's payoff (i.e., from Table 10.1 or Table 10.2) and when information revelation was optional, more dictators chose the "selfish" payoff than when the recipient's payoffs were automatically revealed. Related experiments find that people go out of their way to avoid being asked for a donation, i.e., to be deliberately ignorant of the donation request (Andreoni et al. 2017; DellaVigna et al. 2012).

Freddi (2017) finds that people avoid news articles about a refugee crisis as part of a psychological (intrapersonal) coping strategy to suppress guilt and escape the responsibility of helping to welcome refugees in one's own community. In another example, d'Adda et al. (2018) find that on hot days some people choose to remain deliberately ignorant of the costs of high air-conditioning use so they do not feel pressured to limit their own usage, thus allowing them to express their ignorance if confronted by others.

Grossman and Van Der Weele (2017) propose a signaling model that combines preferences over material payoffs with an intrinsic concern for social welfare and a preference for a self-image as a prosocial actor. They assume that people vary in their degree of prosociality and in the importance they place on self-image. They describe a sequential game as follows:

From "Deliberate Ignorance: Choosing Not to Know," edited by Ralph Hertwig and Christoph Engel. Strüngmann Forum Reports, vol. 29, Julia R. Lupp, series editor. Cambridge, MA: MIT Press. ISBN 9780262045599

- Nature selects the level of social benefit associated with the prosocial action and the individual's type (i.e., how much the individual cares about the social benefit and how much the individual cares about his self-image as a prosocial actor).
- The individual chooses whether or not to receive a signal informing him about the level of social benefit associated with the prosocial action.
- The individual chooses whether or not to take the prosocial action.
- The individual forgets his actual type, goes back to his prior belief about the distribution of types and updates his beliefs about his own type based on the action he took (or did not take) and the signal he got (or did not get) about the action's social benefit.

In the Perfect Bayesian Equilibrium of the game, there are some moderately prosocial individuals who choose not to receive the signal about the action's social benefit and who then choose not to take the prosocial action.

When utility depends on the effects of actions on others, the situation can quickly get very complicated. Thus, a lot of work would be required to build a unified model that successfully incorporates "morality."

Deliberate Ignorance as a Coordination Device

In some cases, deliberate ignorance may be regarded as a way not to undermine a given equilibrium outcome (see also Hoffman et al. 2016). Consider, for example, two countries that face public pressure to intervene in a war zone, should there be evidence that one of the war parties uses chemical warfare, but that acting unilaterally would be insufficient to resolve the conflict. To model such a scenario, suppose there is a first stage in which both countries can look for evidence of chemical weapons. In the subsequent second stage, the two countries decide whether to intervene based on the evidence found in the first stage. Suppose both countries agree on a strategy to intervene if and only if evidence is found, and expected payoffs are given by the matrix shown in Table 10.3. (For simplicity, in the case of acting unilaterally, the benefits of meeting public expectations are assumed to be counteracted by the failure to resolve the conflict, resulting in an overall payoff equivalent to that of having taken no action.) Then each country has an incentive not to report (and in fact to not even look for) evidence of chemical weapons. By deliberately ignoring evidence, they are able to coordinate on an equilibrium they both prefer.

Table 10.3 Expected payoffs to row and column players, respectively, in a coordination game.

	Intervention	No intervention
Intervention	30, 30	0, 0
No intervention	0, 0	50, 50

From "Deliberate Ignorance: Choosing Not to Know," edited by Ralph Hertwig and Christoph Engel. Strüngmann Forum Reports, vol. 29, Julia R. Lupp, series editor. Cambridge, MA: MIT Press. ISBN 9780262045599

Time Limits on Interactions

The prisoner's dilemma is a well-known game in which cooperation does not evolve even though everyone would do better if everyone cooperated. As a result, the repeated prisoner's dilemma (in which the prisoner's dilemma is played by the same players numerous times) has become a common framework for looking at situations in which cooperation will or will not emerge. Even with a fixed limit on the number of rounds that will be played, cooperation can evolve so long as players are sufficiently uncertain about how long another individual will cooperate (Kreps et al. 1982; McNamara et al. 2004). It is better for everyone to *not* know when cooperation will stop, than for everyone to know, or for one individual to know *and* others to know that that individual knows. Consequently, deliberate ignorance can be the best choice even in time-limited repeated prisoner dilemma games.

Term limits for politicians may provide a real-world example: arguably it is better to *not* know whether one will be reelected in order to be able to lay policy foundations on a longer-term basis. Stress-testing of banks is another: it is better for banks to agree beforehand that they will deliberately remain ignorant of which bank is the weakest, as knowledge of which is weakest would likely cause runaway selling of that bank.

However, in all these cases, if it were possible to look at the information privately, an individual could still benefit from it. Thus, there has to be a potential signal to others of seeing the data for deliberate ignorance to make sense as a functionally beneficial strategy.

Bounded Rationality and Self-Deception

Deliberate ignorance can be closely related to self-deception in some cases. Some authors (e.g., Frank 1988; Trivers 2011a) argue that self-deception can be adaptive because that prevents an opponent from reading one's intentions from unintentional cues (caused by bounded rationality or automatic responses, such as blushing). If individuals do not know what they are going to do, then others are unable to infer it from their body language. One can see this phenomenon as an example of deliberate ignorance that is evolutionarily selected because of the strategic advantage it provides.

Fights between male elephant seals may provide an example from animal behavior. Here, advantage is derived if one opponent could not infer another's intention to quit, or is even misled by cues ("if I know that the other will give up after the next *n* strikes, I may continue; otherwise I would give up immediately"). In a human context, consider a poker competition: a weak player might benefit from not looking at their cards before the first round of betting, especially when playing against an opponent well-versed in reading body language.

From "Deliberate Ignorance: Choosing Not to Know," edited by Ralph Hertwig and Christoph Engel. Strüngmann Forum Reports, vol. 29, Julia R. Lupp, series editor. Cambridge, MA: MIT Press. ISBN 9780262045599

Refusing a Free Second Opinion from a Reliable Source

We conclude this section with the example of an individual who refuses a second opinion. This can occur despite the fact that the source of the potential second opinion is trusted as honest and of generally sound judgment.

Imagine, for instance, that Alice is in a position to decide whether Bob will be employed at her company. Before any decision has been announced, one of Alice's trusted friends, Carol, approaches Alice and offers to give her opinion on Bob. Alice thanks Carol for the offer, but turns her down. Why? Because Alice has already formed a very strong judgment about Bob (e.g., his references showed him to be a liar), so Carol's opinion will not influence Alice's decision. Alice may foresee that if Carol had a positive impression of Bob, then having given Alice her opinion, Carol might subsequently feel offended by Alice's decision not to hire Bob. It therefore makes sense for Alice to avoid this scenario by deliberately avoiding whatever information Carol has about Bob.

This is one of several cases presented here that we have left in the form of an intuitive example; it could obviously be abstracted to form a model, with costs and benefits relating to each of the individuals and their actions. The benefit of deliberate ignorance in this case is not about improving one's own actions, nor altering the behavior of others to increase one's reward in the immediate term. Instead, deliberate ignorance acts as a signal to others not to judge one's current actions harshly, and is thus a case of deliberate ignorance being chosen to affect longer-term actions of others under different scenarios that are otherwise unrelated to the situation involving deliberate ignorance.

Deliberate Ignorance through Societal Dynamics

When there are population feedbacks and spillover effects, the frequency of deliberate ignorance in a population will depend on social dynamics as well as individual psychology. Models that only address individual processes, therefore, do not suffice in creating a complete understanding of the phenomenon, nor for planning policy interventions. Here we examine general population models and then turn to a straightforward model that includes deliberate ignorance and spillover effects.

Population-Level Models

Population-level models consider the interactions between many individuals and are mainly concerned with analyzing the consequences of those interactions at the population level. Generally, they look for emergent properties of the system, which can be hard to derive from theory without the model doing the work for us. In particular, population-level models determine the time evolution (or equilibrium states) of certain (population-level) quantities,

From "Deliberate Ignorance: Choosing Not to Know," edited by Ralph Hertwig and Christoph Engel. Strüngmann Forum Reports, vol. 29, Julia R. Lupp, series editor. Cambridge, MA: MIT Press. ISBN 9780262045599

$X = [x_1, \ldots, x_n]$, of the considered system. The formal description of this temporal change in X can occur in some applications through an analytical model (e.g., the Lotka-Volterra models for the interactions between prey and predator species) whereas in others, simulation frameworks are used. Regardless of the framework used, population-level models make explicit assumptions about (a) the size of the population of individuals, (b) the properties of the individuals, (c) the population structure, (d) interaction dynamics leading the update of the variables of interest, and (e) demographic processes.

Let us now consider the case of the diffusion of innovation in a heterogeneous population and explore whether patterns that resemble deliberate ignorance, or ignorance, can emerge as a by-product of the interaction dynamics. We assume a finite population of N heterogeneous individuals. Each individual, i, is characterized by its individual attributes, defined as a vector θ_i of cultural features representing, for instance, a different kind of taste or behavior (Axelrod 1997), and its decision to have adopted the innovation yet or not. Further, social interactions (and consequently information flow) between individuals are represented by networks; that is, collections of nodes represent individuals, and links connecting pairs of nodes represent social relations (e.g., Watts 2002). Additionally, individuals possess a homophilistic bias. Very generally, homophily (in particular "choice" homophily) is the tendency of individuals with similar traits (e.g., physical, cultural, and attitudinal characteristics) to interact with each other more than with people with dissimilar traits (Centola et al. 2007; Lazarsfeld and Merton 1954; McPherson and Smith-Lovin 1987; McPherson et al. 2001). This kind of homophily can be modeled by allowing the social network to evolve as a function of cultural similarities and differences between individuals; for a detailed analysis, see Centola et al. (2007). Depending on the chosen model parameter values, the links between individuals may be arranged in such a way that culturally similar individuals tend to be connected more frequently, forming clusters. Networks of this kind are called correlated networks, and the degree of correlation can be interpreted as a measure of the strength of the homophilistic bias. Now, if an innovation is introduced into such a network (e.g., by a small number of innovators), then depending on the chosen update rule, the adoption dynamic can be very different from well-connected situations. In the extreme, we can imagine that the innovation diffuses only through parts of the population due to individuals only being surrounded by others of their own cluster (i.e., individuals similar to them), while knowledge about the innovation may only be present in another cluster. In other words, the homophilistic bias may cause individuals to choose their neighbors selectively; this, in turn, could result in information being received almost exclusively from individuals who are similar and create a barrier to other sources of useful information present elsewhere in the population.

At first blush, this approach produces results that appear similar to cases where individuals choose to be ignorant. However, there is a difference

From "Deliberate Ignorance: Choosing Not to Know," edited by Ralph Hertwig and Christoph Engel. Strüngmann Forum Reports, vol. 29, Julia R. Lupp, series editor. Cambridge, MA: MIT Press. ISBN 9780262045599

between deliberate actions (with a consequence of ignorance) and *deliberate* ignorance. In this case, although the actions of individuals result in ignorance, it does not benefit them to be ignorant, so they are *not* deliberately ignorant.

A Model of Societal Dynamics Involving Deliberate Ignorance

To provide a minimalistic example of a population model that includes deliberate ignorance as a spillover effect, consider a population in which people can choose whether or not to acquire some information, such as the contents of their Stasi file (see Ellerbrock and Hertwig, this volume). Assume that there are three possible states for an individual: ignorant (I), knowledgeable (K), or deliberately ignorant (D) of the contents of their file. Individuals start off ignorant (I) and sometimes consider looking in their files. As they do so, they consider the advice of another person. Advice from ignorant, knowledgeable, and deliberately ignorant individuals has different effects on the probability that a focal individual chooses to open their file (and become knowledgeable, K) or not (and become deliberately ignorant, D). As the proportions of K and D change, so too do the rates of change because people receive, on average, different advice.

Given this setup, what do you think happens? Does K or D dominate the other, depending on the details? Can D eventually replace K? Will D increase but ultimately die out? Or do K and D tend to coexist?

To answer these questions, we express the above in mathematically precise terms. We can represent this model with three differential equations, one each for I, K, and D. Suppose that individuals of type I consider their files at a rate p, and when considering their file, they first meet another member of the population at random. We assume that, in the absence of advice (i.e., having met another individual of type I), individuals of type I become D with probability r, or K with probability $1-r$. When receiving advice from a K individual, the probabilities are instead $q(D)$ and $1-q(K)$. When receiving advice from a D individual, the probabilities are $s(D)$ and $1-s(K)$. Finally, we allow some rate of population turnover, so that new I individuals appear over time (note that this differs from the real case with Stasi files). This means that at a rate f, K and D individuals leave the population and are replaced by new I individuals. All together, these assumptions imply these three differential equations:

$$\frac{dI}{dt} = \left(K + D\right) f - Ip, \tag{10.15}$$

$$\frac{dD}{dt} = IKpq + IDps + I^2 pr - fd, \tag{10.16}$$

$$\frac{dK}{dt} = IKp\left(1-q\right) + IDp\left(1-s\right) + I^2 p\left(1-r\right) - fK. \tag{10.17}$$

This system has only one interesting steady state, given by

From "Deliberate Ignorance: Choosing Not to Know," edited by Ralph Hertwig and Christoph Engel. Strüngmann Forum Reports, vol. 29, Julia R. Lupp, series editor. Cambridge, MA: MIT Press. ISBN 9780262045599

$$I \to \frac{f}{f+p}, \tag{10.18}$$

$$K \to \frac{p\left(\left(1-r\right)f+\left(1-s\right)p\right)}{\left(f+p\right)\left(f+p\left(1+q-s\right)\right)}, \tag{10.19}$$

$$D \to \frac{\left(pq+fr\right)p}{\left(f+p\right)\left(f+p\left(1+q-s\right)\right)}. \tag{10.20}$$

There is typically coexistence of K and D individuals: K and D at steady state tend to both be greater than zero. In hindsight, this is perhaps obvious. Consider, for example, when $p \to 0.1$, $q \to 0.45$, $r \to 0.4$, $f \to 0.02$, and $s \to 0.7$. Under these parameter values, individuals who do not receive advice tend to open their files 60% of the time. Individuals who encounter K tend also to open their files 55% of the time. But individuals who encounter D open their files only 30% of the time. This results in a steady state with more D than K, even though K initially increases more quickly and a majority (55%) of I individuals who meet K choose also to open their files. Figure 10.2 shows the population dynamics for this example.

This model is perhaps the simplest model that can demonstrate spillover effects of deliberate ignorance, and the simplest model is usually the right place to start. More detail could be incorporated to consider additional effects, such as media amplification or additional population structure. More detailed

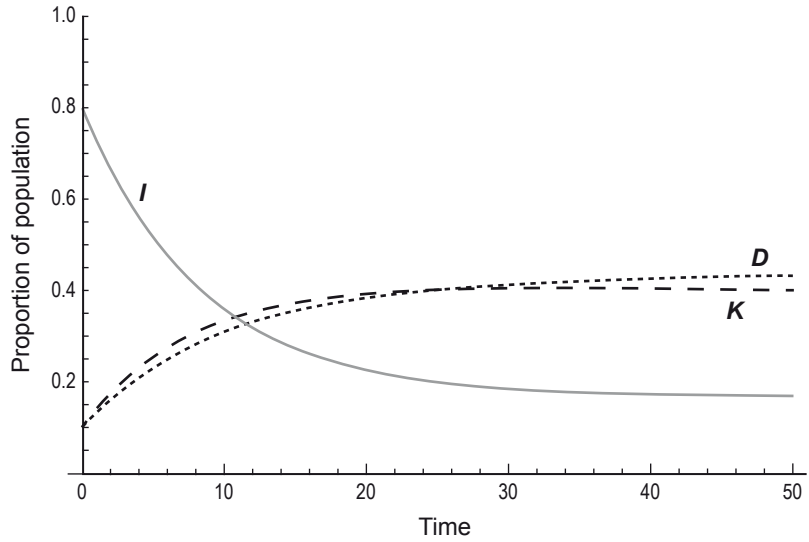

Figure 10.2 After some time, a stable proportion of individuals are deliberately ignorant. Simply ignorant is denoted by I, K stands for knowledgeable about their file, and D signifies deliberately ignorant.

From "Deliberate Ignorance: Choosing Not to Know," edited by Ralph Hertwig and Christoph Engel. Strüngmann Forum Reports, vol. 29, Julia R. Lupp, series editor. Cambridge, MA: MIT Press. ISBN 9780262045599

psychological models could also replace the static *p, q, r, s* parameters, producing more subtle feedback between the population and individual choices.

It is important to note that, while the parameters *q, r, s* embody background knowledge of deliberate ignorance at the psychological level, the population dynamics themselves are quite general. The same equations could apply to many social influence scenarios. This generality of models—abstractions typically apply to more than the contexts that inspire them—is commonplace and can help us appreciate how deliberate ignorance connects to broader phenomena in the study of social dynamics of belief. Quite different psychological mechanisms may produce quite similar population dynamics.

Discussion

In this chapter we have identified several simple theories and models of deliberate ignorance, along with numerous descriptions of situations that could easily be abstracted into models where deliberate ignorance arises. While there are numerous models for individuals, we are aware of very few population-level models where deliberate ignorance emerges in the population through the model dynamics.

We did not discern a general (single) unifying framework for deliberate ignorance, as there can be different fundamental drivers of the phenomenon. Deliberate ignorance can be caused by the expected value of having the information itself (e.g., through automatic mental processes, meaning that such info will likely misguide actions, or for hedonic reasons) or through signaling effects on others (by them knowing that information has/has not been received). We regard this as an important distinction because it is possible for automatic processes (and the like) to evolve into separate systems, and thus for an individual to gain a benefit by no longer being deliberately ignorant. In contrast, when the benefit is one of signaling to others, there can be no such (future) adaptational "improvement."

While some models would say simply *not to bother* collecting, or processing, particular information (e.g., when the cost of acquiring the information is too great), other models identify when it is best to *actively avoid* information. The extent to which we need different models for different types of deliberate ignorance, however, remains an outstanding question.

In real-world situations, there may be multiple causes of deliberate ignorance (e.g., relating to strategic, hedonic, and automatic mental processes). For instance, in the traditional marriage market in India, updated by digital media, parents advertise on marriage websites and construct a consideration set from the responses received. Assume, for instance, that a son's parents are looking for a wife for him. A consideration set may contain between two and five potential wives. The parents then arrange a meeting with the parents of one of the girls, typically at a restaurant, where the two, who may never have met before,

From "Deliberate Ignorance: Choosing Not to Know," edited by Ralph Hertwig and Christoph Engel. Strüngmann Forum Reports, vol. 29, Julia R. Lupp, series editor. Cambridge, MA: MIT Press. ISBN 9780262045599

can talk with each other. Afterward, the parents ask their son whether he agrees with marrying the girl. If he says yes, and the girl also agrees, the search is over. Otherwise, the procedure is repeated with the next girl. The overall situation is very similar to the "secretary problem" in optimal stopping theory. However, this situation is more complex in that rather than the quality of the next potential partner being random, pre-sorting by the parents has occurred before each decision point by the son.

In this process, the young man and the young women are each highly ignorant about the choice set they have. In the extreme, they agree to a future spouse based on a single meeting. Deliberate ignorance enters when young people choose to accept this procedure, rather than searching for themselves. Much of this behavior is hard to understand from the perspective of knowing more is better. However, a large proportion of Indians accept this procedure and have reasons for doing so. Some hold that parents have more experience about what a good spouse is (this relates to whether to use information; their own may be less reliable and mislead them). Some feel that being choosy and rejecting a candidate after a meeting would hurt that person, and thus some do not even want to meet the future wife and simply trust their parents (relating to hedonistic biases). Some view searching themselves as a signal to parents that their judgments are not fully trusted or respected (this strategic aspect may have longer-term ramifications). In addition, the social norms that have developed in that society (arguably driven initially by the previous three aspects) alters the payoff structures associated with such actions (greater consternation on the part of parents, more likely to be judged by friends, and so on).

The information gap belief-based utility model and the optimism model both predict that the choice to remain deliberately ignorant depends on affect (i.e., on how good or bad the beliefs would make a person feel). Different emotions can be similar in affect, yet different on other dimensions. For instance, sadness, fear, and disgust all produce negative affect but are very different from each other. Whether, empirically, people exhibit the same pattern of deliberate ignorance across beliefs that induce different emotions with similarly negative affect is an open question in need of further study.

The idea that it can be better in some cases to mask bad experiences, rather than to hold on to accurate memories of the past, seems relevant to deliberate ignorance, although in this case it is about becoming ignorant of one's own prior experiences. This is potentially very important, given the impact of some conditions (e.g., posttraumatic stress disorder) on individuals. It may be beneficial to have better models relating to this idea in the future.

We also note that the majority of our discussion focused on whether to obtain information in particular settings, rather than the conditions under which mental processing (or storage) of such information would not have evolved. For models relating to when learning is unlikely to evolve, despite being beneficial, see Trimmer and Houston (2014).

From "Deliberate Ignorance: Choosing Not to Know," edited by Ralph Hertwig and Christoph Engel. Strüngmann Forum Reports, vol. 29, Julia R. Lupp, series editor. Cambridge, MA: MIT Press. ISBN 9780262045599

Finally, we note that discussion of what constitutes "deliberate" ignorance can be entertainingly problematic. Consider, for example, a plant that benefits from bet hedging with its seeds, relative to conditions for when to germinate. Suppose that to germinate at different times, some seeds have wide pores (which readily respond to rain by germinating) or small pores (thus being more likely to wait). Does that constitute deliberate ignorance? The seed's shell is stopping rain "information" from triggering it, so its structure is keeping it ignorant of conditions. Arguably, this is *not* deliberate ignorance as the seed itself is not making that choice. Now, what if an animal hedges its bets by producing offspring who differ in whether (or how often) they accept or avoid freely available information? This surely seems like deliberate ignorance when that individual is tested, but the offspring still have had that imposed upon them, just like the seeds. Further, what if during development, an individual had the choice of which type of mental mechanisms to produce. One set of mechanisms would be more accurate if the environment changed (but is more accurate than by having actively ignored the initial conditions); another set would do best under current conditions (by immediately absorbing information about the environment). Is the individual who chooses the set that will subsequently ignore that information being deliberately ignorant? It would certainly seem so. But what if their probability of choosing that set were already genetically set for them? One perspective is that for something to be "deliberate," some cost must be imposed by the deliberative process, as the action (or in this case, ignorance) may otherwise occur without being deliberative. Ultimately though, agents perform actions, and natural selection then acts without any necessary distinction of what is, or is not, "deliberate." What constitutes "deliberate ignorance" may therefore always be blurry around the edges, when real biological systems are addressed.

Acknowledgments

Particular thanks to Ulrike Hahn, Ralph Hertwig, Simon Gächter, Kristin Hagel, Stephan Lewandowsky, Pete Richerson, and Barry Schwartz. PCT was also partly funded by the German Research Foundation (DFG) as part of the SFB TRR 212 (NC^3).

From "Deliberate Ignorance: Choosing Not to Know," edited by Ralph Hertwig and Christoph Engel. Strüngmann Forum Reports, vol. 29, Julia R. Lupp, series editor. Cambridge, MA: MIT Press. ISBN 9780262045599

From “Deliberate Ignorance: Choosing Not to Know,” edited by Ralph Hertwig and Christoph Engel.
Strüngmann Forum Reports, vol. 29, Julia R. Lupp, series editor. Cambridge, MA: MIT Press. ISBN 9780262045599