DIVISIBLE UPDATING

MARTIN W. CRIPPS

Abstract. A characterisation is provided of the belief updating processes that are independent of how an individual chooses to divide up/partition the statistical information they use in their updating. These "divisible" updating processes are in general not Bayesian, but can be interpreted as a re-parameterisation of Bayesian updating. This class of rules incorporates over- and under-reaction to new information in the updating and other biases. We also show that a martingale property is, then, sufficient the updating process to be Bayesian.

1. Belief Updating

In this paper we consider arbitrary processes for updating beliefs in the light of statistical evidence. We treat these updating processes as deterministic maps from the prior beliefs and a signal structure (statistical experiments) to a profile of updated beliefs (one for each possible signal). We place axioms on these updating processes and show that a generalisation of Bayesian updating can be derived as a consequence of these axioms.

Consider an individual who receives two signals/pieces of information/news. There are several ways such an individual can use these two signals to update their beliefs about the world. One is to consider the joint distribution of these signals and to do just one update. An alternative (which is natural when the signals arrive sequentially or over time) is to separate the two signals and to update beliefs twice. That is, to update beliefs once using the first signal and its distribution. And then to update these intermediate beliefs a second time using the second signal and its conditional distribution given the realisation of the first signal. If these two different procedures generate the same ultimate profile of updated beliefs we will say that the updating is divisible. We will characterise all the updating processes that have this divisibility property and show that they can be interpreted as natural generalisation of Bayesian updating. Furthermore, we will show that divisibility plus an unbiasedness/martingale property for the updating characterises Bayesian updating.

Date: December 11, 2018.

Department of Economics, University College London; m.cripps@ucl.ac.uk. My thanks are due to Aislinn Bohren, Philippe Jeheil and Ran Speigler for their suggestions and comments and also to seminar audiences at Caltech, MIT-Harvard, PSE and Penn.

The main result of this paper shows that if updating satisfies four properties, then it is characterised by a homeomorphism F from the space of beliefs into itself. The four properties or axioms are: First, that uninformative experiments do not result in changes in beliefs. Second, that the names of the signals do not matter just their probability content. The third ensures that all updated beliefs are possible. The fourth is the divisibility property described above. Any updating procedure, that satisfies these four properties, follows the steps that are illustrated in the figure below. The initial beliefs are mapped to a "Shadow Prior" using a homeomorphism F. Then, these shadow priors are updated in the standard Bayesian fashion using the statistical information that is observed to create a Shadow Posterior. Finally, the shadow posterior is mapped back to the space of original beliefs using the inverse map F^{-1}.

FIGURE 1. Updating that Satisfies Divisibility.

Our result is that if the belief revision protocol satisfies the divisibility property, then it must have this structure. It is clear that Bayesian updating is a subset of this class, (F is the identity). Specific properties of F ensure different properties of the belief revision. Biases in the updating will be generated by the convexity or concavity of the map F. For overreaction and under-reaction to signals we will argue that F needs to be an expansive or a contraction mapping. In regions where F is a contraction the belief revision will exhibit over-reaction in general and in regions where F^{-1} is a contraction the belief revision will exhibit underreaction.

Many of the useful properties of Bayesian updating also carry over to this class of revision protocols. For example, although the actual beliefs are not a martingale the shadow beliefs are. Thus, in dynamic settings, updating that satisfies divisibility will be a deterministic function of a martingale. As a result, consistency will hold for this larger class of updating processes, provided F maps the extreme points of the belief simplex to themselves. That is, when these updating processes are repeatedly applied to a sequence of data, they will generate limiting beliefs that attach probability one to the truth. This property of divisible updating contrasts with other models of non-Bayesian updating, for example Rabin and Schrag (1999) and Epstein, Noor, and Sandroni (2010).

A question we also address is: what additional axioms are necessary to characterise Bayesian updating? We show in the final section of this paper that only one further property is sufficient for this, that is, belief revision is unbiased or follow a martingale.[1] Thus

[1]There are several names used for this notion in different contexts: unbiased, Bayes plausibility have also been used.

any relaxation of Bayesian updating has to either violate the martingale property, the divisibility property or one of the other axioms we impose.

Why do we focus on the divisibility of updating rather than some other property it may have? There are several motivations for focussing on divisibility. The first is normative. This is a property that is readily understandable by subjects and is a property that they would like their updating to have. This is in contrast to the martingale property of Bayesian updating (that the expectation of future beliefs equals the current belief), which is both difficult to explain and harder to motivate as a property of updating.

The first is it resolves simplifies the determination of beliefs. The information economic agents observe is typically not a unitary but a bundle of signals. It may be that even apparently simple pieces of information (like the temperature is -5^oC) is parsed into several separate signals: “it is below freezing”, “it is also -5^oC”. What divisibility ensures is that the exact way that the updating is performed (using any bundle of signals) is unimportant in determining the final beliefs. So, if the individual separates out various features of the bundle and uses these individually to update their beliefs, then the order that they do this has no ultimate effect their eventual beliefs. If updating does not satisfy the divisibility property, then the updated beliefs will depend on how and in what order the individual has applied the updating (see Section 1.1 below for an example of this).

The second is modelling parsimony. If divisibility is not satisfied, the individual has (in general) multiple possible updated beliefs and an additional assumption is required to model them. To predict the individual’s updated beliefs it will, then, also be necessary to have a theory of how she chose to apply the updating rules to packages of signals.

The third is that treating the agent’s current beliefs as a state variable in their decision taking is no longer legitimate. Agents may benefit by recalling more of the history than just her current state of beliefs if the order of the updating matters. In dynamic settings this issue is most clear. If the updating satisfies divisibility, then it is only necessary for the individual to keep track of their current beliefs and update them any time new information appears. Her current state of beliefs are sufficient to summarise all past signals in a parsimonious way—nothing else about the past history must be remembered. There is a dynamic consistency in the updating so the individuals who are not required to act can simply collect information and use it to update beliefs when it is necessary. If the updating is not divisible, then how often the updating rule has been applied in the past may affect her current beliefs and thus may be something she needs to keep track of. In summary, if the belief updating is not divisible, then a theory of how and when updating occurs is required. Such a theory needs to address the trade-offs between the memory costs of storing accumulated signals and the processing costs of when and how to update.

1.1. *An Example of Non-Divisible Updating*

In this section we consider the model of non-Bayesian learning described Epstein, Noor, and Sandroni (2010) or conservative Bayesian updating of Hagmann and Loewenstein (2017) and apply it to an example.[2] In these models of learning, the updated beliefs are a linear combination of the prior and the posterior. We will show, using an example, that this model of updating does not satisfy the divisibility property and seek to explain why the updated beliefs vary for different protocols.

Consider an individual who is waiting for a bus and is learning about the arrival process of buses. There are two states for the arrival process: In the good state a bus will arrive in period $t = 0, 1, \ldots$ with probability $(1-\alpha)\alpha^t$ ($\alpha \in (0,1)$) while in the bad state a bus arrives in period t with probability $(1-\beta)\beta^t$ where $\alpha < \beta$. She has initial beliefs $\mu \in (0,1)$ that the state is good and a bus will arrive at some point. If no bus arrives in period $t = 0$, then a Bayesian would revise these beliefs downward in the light of this bad news to $\frac{\alpha\mu}{(1-\mu)\beta+\mu\alpha}$. However, in this model of non-Bayesian learning she revises her beliefs to a weighted average of the prior and Bayesian posterior:

$$\mu_1 = (1-\lambda)\mu + \lambda \frac{\alpha\mu}{(1-\mu)\beta + \mu\alpha}, \qquad \lambda \geq 0. \tag{1}$$

This is a particularly elegant model as it preserves the martingale property of the Bayesian posteriors. The value λ is a parameter of the updating procedure, which can in general depend on the current history, but in this example it is treated as a fixed property of the updating. This model of updating generalises Bayesian updating ($\lambda = 1$), where λ adjusts the effect the bad news of no arrival has the beliefs. Choosing $\lambda < 1$ allows the individual to be under-confident or conservative in their updating of their beliefs. Conversely, $\lambda > 1$ allows the individual to be overconfident about the new information they have received. Thus, λ allows a range of updates from extreme dogmatism ($\lambda = 0$) or jumping to certainty ($\lambda \to \mu^{-1}(1-\alpha)^{-1}$).

Now consider what the individual's beliefs could be after t periods waiting for a bus without an arrival. One possible application of this updating procedure is to iterate the updating process described in (1) for each of the t periods the individual has been waiting, to get

$$\mu_\tau = (1-\lambda)\mu_{\tau-1} + \lambda \frac{\alpha\mu_{\tau-1}}{(1-\mu_{\tau-1})\beta + \mu\alpha}, \qquad \tau = 0, 1, \ldots, t.$$

This updated belief in period t, μ_t, clearly has required considerable mental agility on the part of the individual. But it requires less memory, as the individual does not need to keep track of all past events—the current value of the beliefs μ_s is all that she needs to know.

[2] Both of these papers use this model of updating as a part of more general study of dynamic choice when there is learning or information design. We do not address choice issues here, focussing solely on belief revision.

An alternative application of the updating procedure (1) would be to suppose that the individual took all their current information at time t and performed one update of their prior. This is the update (for example) that is done by someone who only gets to the bus stop in period t and learns a bus has not yet arrived. Such a person would do one Bayesian update and arrive at the updated belief $\frac{\alpha^t \mu}{(1-\mu)\beta^t + \mu\alpha^t}$.[3] Hence the non-Bayesian updating protocol (1) would in this case have an updated belief:

$$\tilde{\mu}_t := (1-\lambda)\mu + \lambda \frac{\alpha^t \mu}{(1-\mu)\beta^t + \mu\alpha^t}.$$

The updated belief $\tilde{\mu}_t$ has required a different kind of mental agility—requiring the individual to keep track of the past history of events and their original prior. The feature of the updating procedure (1) we seek to remedy below is that the iterated and one-shot update result in different beliefs: $\tilde{\mu}_t \neq \mu_t$. The two different ways of using the same information has generated different updated beliefs. The only time $\tilde{\mu}_t = \mu_t$ is when $\lambda = 1$ and the updating is fully Bayesian.

There is no simple comparative static that describes when one of the updating procedures described above makes the individual more optimistic than the other. If we take the initial beliefs $\mu = \frac{1}{2}$ for simplicity, then $\tilde{\mu}_2 = (1-\lambda)\frac{1}{2} + \lambda\frac{\alpha^2}{\beta^2+\alpha^2}$ and the iterated updating gives: $\mu_1 = (1-\lambda)\frac{1}{2} + \lambda\frac{\alpha}{\beta+\alpha}$, and $\mu_2 = (1-\lambda)\mu_1 + \lambda\frac{\alpha\mu_1}{(1-\mu_1)\beta+\mu_1\alpha}$. A calculation (see Appendix) shows that

$$\frac{\tilde{\mu}_2 - \mu_2}{\lambda\gamma(1-\gamma)} = \frac{1}{\frac{1}{\mu_1} + \frac{\gamma}{1-\mu_1}} - \frac{1}{1+\gamma+\gamma^2+\gamma^3}, \qquad \gamma := \frac{\alpha}{\beta}, \quad \mu_1 \in (0, 1/2).$$

This implies that $\tilde{\mu}_2 > \mu_2$ if and only if $\lambda < 1$. The intuition for this comparative static is as follows. Varying λ tends to emphasise either the data or the prior. The iteration of the updating mitigates this effect. To see how this works in practice first suppose that $\lambda < 1$ and the prior is given increased weight. In the one-shot update, $\tilde{\mu}_2$, there is only one opportunity for the bad news—two periods without busses—to drive the prior down. But, when updating is iterated, it weakens the effect of the increased weight on the original prior. It, instead, places some weight on the intermediate prior, μ_1. Hence, μ_2 will tend to be smaller than $\tilde{\mu}_2$ when the information that arrives is bad news, because μ_2 places less weight on the original prior than $\tilde{\mu}_2$.

Conversely, when $\lambda > 1$ the data is given increased weight in the updating. The one-shot update, $\tilde{\mu}_2$, gives these two periods of bad news excessive weight. Whereas the iterated update, μ_2, decreases the weight given to the first period of the bad news by averaging it with the prior. So, the individual is less optimistic when they do a one-shot up date because it maximises the weight given to the data that has been collected. The iteration now dilutes the effect of the over-emphasised data.[4]

[3] α^t is the probability that no bus arrives in the first t periods in the good state.

[4] When t varies continuously it is possible to get a cleaner expression for μ_t and a similar comparative static on μ_t and $\tilde{\mu}_t$ can be obtained.

Of course, the two ways of determining period t beliefs, μ_t and $\tilde{\mu}_t$, are not the only ways the beliefs in period t could be arrived at. An individual could update several times over the periods $s = 0, \dots, t-1$—how many times this occurred would depend on the agent's costs of cogitation. Thus there is a whole family of potential updated beliefs at time t and as time passes this family grows.

1.2. *Examples of Divisible Updating*

Here we describe three other classes of updating that do satisfy the divisibility property and are non-Bayesian. These rules will only be stated in terms of the above example but are derived from completely general updating functions that will be developed later in the paper. The first is a simple generalisation of Bayes' rule

$$\mu_1 = \frac{\alpha^x \mu}{\mu\alpha^x + (1-\mu)\beta^x}, \qquad x \geq 0.$$

A second would move the inverse probabilities apart by a factor that depends on the information in the signals

$$\frac{1}{\mu_1} - \frac{1}{1-\mu_1} = \frac{1}{\mu} - \frac{1}{1-\mu} + y \ln \frac{\beta}{\alpha}.$$

A final one would use trigonometric functions

$$\mu_1 = \frac{2}{\pi} \arctan\left(\frac{\sqrt{\alpha}}{\sqrt{\beta}} \tan \frac{\pi}{2}\mu\right).$$

To verify that these updating rule satisfy divisibility is a trivial exercise. What is surprising about the final updating rule is that there is no role for $1-\mu$ in the updating function. This follows because of the identity $\cot\theta = \tan(\frac{\pi}{2} - \theta)$.

1.3. *Related Literature*

There is a growing Economics literature, both experimental and theoretic, investigating the consequences of a non-Bayesian updating of beliefs, see for example: Rabin and Schrag (1999), Ortoleva (2012), Angrisani, Guarino, Jehiel, and Kitagawa (2017), Levy and Razin (2017), Brunnermeier (2009), Bohren and Hauser (2017) among many others. Much of this literature combines issues of updating and decision taking. This is not what the current paper does—it focusses solely on the revision of beliefs and the properties one might want to place on this revision. One theme of this literature has been to investigate the properties of a particular assumption about how updating may occur. The aim here is somewhat different, that is to try to understand what updating procedures are consistent with a given property. One exception to the focus on decision taking is Epstein, Noor, and Sandroni (2010) who provide a model of updating that captures the under and overreaction to new information. Their model of updating is distinct from the class we consider in several respects and has already been considered at length in the example above.

In Gilboa and Schmeidler (1993) the notion of divisibility (termed there commutativity) is introduced and is argued to be an important feature of belief updating particularly in the context of updating ambiguous beliefs. Hanany and Klibanoff (2009), has perhaps the closest connection with this one. Here it is shown that there is a unique "reweighted Bayesian update" that generates a given set of dynamically consistent preferences. They moreover show that this rule satisfies commutativity, a property equivalent to divisibility. These reweighted Bayesian updates are a subset of the class of updating rules that described here (the rules that update as in Figure 1). Here we described all continuous updating rules that satisfy commutativity/divisibility. In Zhao (2016) a set of weaker axioms are shown to characterise an updating rule that does not satisfy divisibility, but does satisfy an order independence property similar to the one discussed in the appendix, however, this property is required to hold only for independent events.

There are links between the notion of divisibility and the models dynamic choice under uncertainty. In particular the literature on recursive preferences in dynamic settings, or dynamically consistent preference update rules. Here agents are required to act consistently in situations where information arrives over many periods and thus implicitly behave as if they update divisibly; see for example Epstein and Zin (1989) or Epstein and Schneider (2003)).

In the statistics literature the notion of prequentiality, Dawid (1984), emphasises the idea that that forecasting should be an iterative procedure and there may be differences between iteratively revised forecasts and other kinds of forecasting procedures.

2. A Model of Belief Updating

In this section we describe our model of belief revision and the most important axioms we will impose on the updating. The approach taken is inspired by the axiomatic interpretation of entropies: see for example Shannon and Weaver (1949),Tverberg (1958) or Aczél and Dhombres (1989) p.66. We will not adopt the terminology of "priors" and "posteriors", reserving these terms for the Bayesian updating only. Instead the agent is assumed to be equipped with "beliefs" that are revised when information arrives to form "updated beliefs".

There is an unknown parameter $\theta \in \{1, 2, \ldots, |\Theta|\} := \Theta$ and an agent who has the initial beliefs, $\mu = (\mu^1, \ldots, \mu^{|\Theta|}) \in \Delta(\Theta)$, about the value of this parameter. There is a statistical experiment $\mathcal{E}$ that the agent can conduct which provides further information on θ.[5] The experiment comprises a finite set of signals $s \in \{1, 2, \ldots, n\} = S$ and parameter-dependent probabilities for the signals $p^\theta = (p^\theta_1, \ldots, p^\theta_n) \in \Delta(S)$. We will only consider experiments with strictly positive probabilities for all signals: $p^\theta_s > 0$ for all $s \in S$ and $\theta \in \Theta$, or that $p^\theta \in \Delta^o(S)$.[6] We will also want to consider the probabilities of a given signal $s \in S$ as the

[5]See Torgersen (1991), for example, for the general properties of statistical experiments.

[6]We use $\Delta^o(S)$ to denote the interior of $\Delta(S)$.

parameter varies, hence we will define $p_s := (p_s^1, \dots, p_s^{|\Theta|}) \in (0,1)^{|\Theta|}$. Thus p^θ are the rows of the matrix $\mathcal{E}$ and p_s are its columns. In summary, the agent has priors $\mu \in \Delta(\Theta)$ and access to the experiment $\mathcal{E} := (p^\theta)_{\theta\in\Theta} \in \Delta^o(S)^{|\Theta|}$.

An updating process takes the experiment, $\mathcal{E}$, and the beliefs, μ, and maps them to the updated beliefs for each possible signal outcome, $s \in \{1, 2, \dots, n\}$. The outcome of the updating process is a profile of n possible updated beliefs $\{\mu^s \in \Delta(\Theta) : s \in S\}$ (one for each possible signal realisation $s \in S$). The "updating function" $\mathcal{U}_n$ is defined to be the map from the beliefs and the experiment to the profile of updated beliefs, that is

$$\mathcal{U}_n : \Delta(\Theta) \times \Delta^o(S)^{|\Theta|} \to \Delta(\Theta)^n, \qquad \text{for } n = 2, 3, \dots \ .$$

We will also write $(\mu^1, \dots, \mu^n) = \mathcal{U}_n(\mu, (p^\theta)_{\theta\in\Theta}) = \mathcal{U}_n(\mu, \mathcal{E})$.[7]

The first condition the updating is required to satisfy is that if the signals are identical then there is no updating. That is if $p^\theta = p^{\theta'}$ for all $\theta, \theta' \in \Theta$ then the updated belief is the same as the original beliefs.

Axiom 1 (Uninformativeness)**.** $\mathcal{U}_n(\mu, \mathcal{E}) = (\mu, \dots, \mu)$, if $p^\theta = p^{\theta'}$ for all $\theta, \theta' \in \Theta$.

The second condition is that the names of the signals are unimportant for how the beliefs are revised, it is only the probabilities in the experiment that matter. Thus permuting the order of the signals just permutes the order of the updated probabilities.

Axiom 2 (Symmetry)**.** For any n, any permutation $\omega : \{1, 2, \dots, n\} \to \{1, 2, \dots, n\}$, any μ, and any $\mathcal{E} = (p^\theta)_{\theta\in\Theta} \in \Delta^o(S)^{|\Theta|}$,

$$\mathcal{U}_n(\mu, (\omega(p^\theta))_{\theta\in\Theta}) = \left(\mathcal{U}_n^{\omega(1)}(\mu, \mathcal{E}), \dots, \mathcal{U}_n^{\omega(n)}(\mu, \mathcal{E})\right),$$

where $\omega(p^\theta) := (p^\theta_{\omega(1)}, \dots, p^\theta_{\omega(n)})$ and $\mathcal{U}_n(.) = (\mathcal{U}_n^1(.), \dots, \mathcal{U}_n^n(.))$.

The final condition in this section requires that the updating is non-dogmatic and continuous. Continuity is satisfied by many models of belief revision in the literature, but may not hold if there are fixed costs of contemplation and belief revision; see Ortoleva (2012) for an example of discontinuous updating. Non-dogmatic revision, loosely stated, allows an agent to have arbitrary updated beliefs after a signal s if they observe sufficiently persuasive evidence. To be more precise, it says that if there are only two possible signals, then for any initial belief (that attaches positive probability to every parameter) there exists a unique experiment that generates an arbitrary updated belief after signal $s = 1$. This does *not* require that all possible *profiles* of beliefs can be generated from a suitable statistical experiment. It only requires that the range of updating function for a given signal is the entire set $\Delta^o(\Theta)$. The additional requirement that this map is a bijection is required for complete solution of the functional equation we later solve. Without uniqueness only local

[7]In this model the updated beliefs are a deterministic function of the signal and experiment. This is not consistent with all models of updating. For example in Rabin and Schrag (1999) the updated beliefs are randomly determined by a bias that is realised after the signal is observed. To capture this model of updating it would be necessary for the function U_n to take values in $\Delta(\Delta(\Theta))^n$.

solutions for U_n exist.[8] The uniqueness property would again be violated by models updating that have fixed costs of contemplation. In such models there may be sets of experiments for which it is simply not worth revising beliefs, hence there would be many experiments with the update equal to the prior.

Axiom 3 (Non-Dogmatic)**.** The function $\mathcal{U}_2 : \Delta(\Theta) \times \Delta^o(S)^{|\Theta|} \to \Delta(\Theta)^2$ is continuous. For any $\mu, \mu' \in \Delta^o(\Theta)$ there exists a unique $\tilde{\mathcal{E}} \in \Delta^o(S)^{|\Theta|}$ such that $\mathcal{U}_2^1(\mu, \tilde{\mathcal{E}}) = \mu'$.

2.1. Binary Experiments

We end this section by introducing the function u that describes the updating in the case where there are only two signals—a binary experiment.[9] This will play a key role in what follows. Consider an experiment with only two signals: $S = \{1, 2\}$ and $\mathcal{E} = (p_1^\theta, p_2^\theta)_{\theta\in\Theta}$. We can write

$$\mathcal{U}_2(\mu, \mathcal{E}) \equiv \left(\mathcal{U}_2^1(\mu, p_1, p_2), \mathcal{U}_2^2(\mu, p_1, p_2)\right).$$

Recall that p_s is the vector of parameter dependent probabilities for signal $s = \{1, 2\}$, so $p_2 = \mathbb{1} - p_2$. The axiom of symmetry implies that the two functions on the right above are identical when the arguments are transposed, that is, $\mathcal{U}_2^1(\mu, p_1, p_2) \equiv \mathcal{U}_2^2(\mu, p_2, p_1)$. So, we define $u : \Delta(\Theta) \times (0, 1)^{|\Theta|} \to \Delta(\Theta)$ as

$$u(\mu, p_1) := \mathcal{U}_2^1(\mu, p_1, \mathbb{1} - p_1). \tag{2}$$

That is, $u(\mu, p_1)$ describes the updated beliefs after a binary experiment where the signal $s = 1$ occurred and $p_1 \in (0, 1)^{|\Theta|}$ were the parameter-dependent probabilities of the signal $s = 1$ and $\mathbf{1} - p_1$ were the probabilities for $s = 2$. This allows us to write the full profile of updated beliefs for binary experiments only in terms of one function u: and

$$\mathcal{U}_2(\mu, \mathcal{E}) \equiv (u(\mu, p_1), u(\mu, p_2))$$

We will use divisibility to extend this decomposition so that it holds for experiments with arbitrary numbers of signals (4). Axiom 1 applied to a binary experiment implies that the function $u(.)$ satisfies

$$u(\mu, p\mathbb{1}) = \mu, \qquad \forall \mu \in \Delta(\Theta), p \in (0, 1). \tag{3}$$

---

[8]See, for example, Berg (1993).

[9]A dichotomy is a experiment where there are only two parameters.

3. Axioms on Iterated Updating: Divisibility and Order Independence

We now consider they key condition on the updating of beliefs that is studied in this paper. The Axiom 4 is designed to capture the fact that the update is independent of whether information is processed in an iterative way or as one-off process.[10]

Consider two possible ways of learning the signal s. The first is a one-step process where a signal s is generated according to the experiment $\mathcal{E} = (p^\theta)_{\theta\in\Theta}$ and then revealed to the individual. The second is a two-step process where, first, with probabilities $(p_1^\theta, 1-p_1^\theta)_{\theta\in\Theta}$ the signal $s = 1$ or the signal $s \neq 1$ is revealed to the agent in a simpler experiment. Then, in the case where the outcome $s \neq 1$ was obtained in the first experiment, a signal from the set $\{2, \dots, n\}$ is generated from a second experiment with the correct conditional probabilities $(p_{-1}^\theta(1-p_1^\theta)^{-1})_{\theta\in\Theta}$ (where p_{-s}^θ is the vector p^θ with the s^{th} element omitted). In what follows we will use $\mathcal{E}_{-1} := (p_{-1}^\theta(1-p_1^\theta)^{-1})_{\theta\in\Theta}$ to denote the experiment that occurs conditional on $s \neq 1$.[11]

The Axiom 4 says that these two different processes for observing the signal s have no effect on the individual's ultimate profile of beliefs. This assertion has two distinct elements. First it says that learning the signal is $s = 1$ when there are $n-1$ other signals has the same effect on the updated beliefs as learning the signal is $s = 1$ when there is a binary experiment and the other signal occurs with probabilities $1-p_1^\theta$. Formally this requires that $\mathcal{U}_n^1(\mu, \mathcal{E}) \equiv u(\mu, p_1)$. Thus the relative probabilities of the signals that were not observed have no role in determining how beliefs will be updated.

The second element of the Axiom 4 says that the updated beliefs an individual has when they see the signal $s' > 1$ in a one-off experiment ($\mathcal{U}_n^{s'}(\mu, \mathcal{E})$) are the same as the beliefs they would have at the end of the two-step process. In this two step process they first learn the signal was not $s = 1$ and updated their beliefs to $u(\mu, \mathbf{1} - p_1)$ and then they learn that the signal was s' from the experiment $\mathcal{E}_{-1}$ and engaged in the further update to $\mathcal{U}_{n-1}^{s'-1}(u(\mu, \mathbf{1} - p_1), \mathcal{E}_{-1})$.[12] [13]

Axiom 4 (Divisibility)**.** For all $\mathcal{E} = (p^\theta)_{\theta\in\Theta} \in \Delta^o(S)^{|\Theta|}$, $\mu \in \Delta(\Theta)$, and $n \geq 3$

$$\mathcal{U}_n(\mu, \mathcal{E}) = [u(\mu, p_1), \mathcal{U}_{n-1}(u(\mu, \mathbf{1} - p_1), \mathcal{E}_{-1})] .$$

[10] As an alternative, in the Appendix we consider, Axiom 6, which requires that the updating of beliefs is independent of the order that information arrives. (Reversing the order that two pieces of information arrive has no effect on the ultimate profile of beliefs.) Although these axioms appear different, we show in the Appendix that they are actually equivalent and so we will only use the divisibility axiom.

[11] Recall that $p^\theta \in \Delta^o(\Theta)$ and so $1 > p_1^\theta$.

[12] We use $\mathbf{1}$ to denote the vector $(1, 1, \dots, 1)$ of appropriate length.

[13] The form of this function requires a little explanation. The updated beliefs after the signal s' is the $s'-1^{\text{th}}$ component of the profile $\mathcal{U}_{n-1}$ when the first signal is not present. Hence the change in the superscript on $\mathcal{U}_{n-1}$.

If this axiom is combined with symmetry and iteratively applied, it implies that any multi-step procedure that is based upon the experiment $\mathcal{E}$ will result in the same updated profile of beliefs.

4. Characterisation of Divisible Updating

In this section a proposition is proved that gives a full characterisation of any updating procedure U_n that satisfies the Axioms 1, 2, 3, and 4. We will show that any such updating is characterised by a homeomorphism F that maps beliefs to shadow prior. Then, this shadow prior is updated in a fully Bayesian manner to create a shadow posterior. Finally the shadow posterior is mapped back by F^{-1} to form the individual's updated beliefs. This is illustrated in the figure below.

$$\begin{array}{ccc} \text{beliefs} & \xrightarrow{\quad U \quad} & \text{updated beliefs} \\ \big\downarrow F & & \big\uparrow F^{-1} \\ \text{shadow prior} & \xrightarrow{\text{Bayes Updating}} & \text{shadow posterior} \end{array}$$

Figure 2. Updating Functions $\mathcal{U}_n$ that Satisfy Axioms 1, 2, 3, and 4.

The first step in the argument is to show that $u(\mu, p_s)$, the function that determines the updated beliefs after signal s, must be homogeneous degree zero in p_s if it satisfies Axioms 1, 2, and 4. The intuition for this result is quite simple. Suppose that there are three possible signals $S = \{1, 2, 3\}$ and that one signal, say $s = 2$, is equally likely under all states. We now apply the Axiom 4 to the two cases: one where s is determined in a one-off experiment and the second where there is first an experiment with binary outcomes $s = 2$ and $s \neq 2$ and then $s \in \{1, 3\}$ is selected with the appropriate conditional probabilities. As observing $s = 2$ is uninformative, Axiom 1 implies the first stage of the two-step process leads to no updating of the priors. However the second stage experiment, when a signal $s \neq 2$ is selected, has increased the relative probabilities p_s of the signals $s \in \{1, 3\}$. Thus the equality in Axiom 4 implies that the beliefs after the one-step experiment (with low relative probabilities p_s of the signals $s \in \{1, 3\}$) are equal to the beliefs after the two-step experiment (with higher relative probabilities of the signals). Thus scaling up the probabilities had no effect on the updating of beliefs. This is the definition of homogeneity degree zero.

Lemma 1. *Suppose updating satisfies Axioms 1, 2, and 4, then the function $u(\mu, p_s)$ defined in (2) is homogeneous degree zero in p_s. And*

$$\mathcal{U}_n(\mu, \mathcal{E}) \equiv (u(\mu, p_1), \dots, u(\mu, p_n)) \,. \tag{4}$$

Proof. Suppose that $n = 3$. Consider two ways the agent can process the signal $s = 1$: (a) She could be told the outcome of an $s = 1$ or $s > 1$ experiment, which would result in the updated beliefs (5). (b) The agent could first be told $s \neq 2$ in an $s = 2$, $s \neq 2$ experiment and then the outcome of a final experiment where $s \in \{1, 3\}$, which would result in the two step second updating (6).

$$u(\mu, p_1) \tag{5}$$

$$u\left(u\left(\mu, \mathbf{1} - p_2\right), \left(\tfrac{p_1^\theta}{1-p_2^\theta}\right)_{\theta\in\Theta}\right) \tag{6}$$

The combination of Axioms 2 and 4 applied to (5) and (6) implies

$$u(\mu, p_1) \equiv u\left(u\left(\mu, \mathbf{1} - p_2\right), \left(\tfrac{p_1^\theta}{1-p_2^\theta}\right)_{\theta\in\Theta}\right) \qquad \forall p_2^\theta \in [0, 1 - p_1^\theta]. \tag{7}$$

If $1 - p_2^\theta = 1/\lambda$ for all θ, then (3) implies $u\left(\mu, \mathbf{1} - p_2\right) \equiv \mu$. So the right of (7) becomes $u(\mu, (\lambda p_1^\theta)_{\theta\in\Theta})$ and we get the condition

$$u(\mu, p_1) \equiv u\left(\mu, \lambda p_1\right), \qquad \forall \lambda \in [1, \min_\theta (p_1^\theta)^{-1}].$$

Hence the function u is homogeneous degree zero in p_1 if Axiom 4 holds. The equation (4) follows by combining Axiom 4 with symmetry. □

We now move on to establishing the main result which gives the form of the functions $u(\mu, p^s)$ defined in (2). The intuition for this characterisation of the function $u(\mu, p^s)$ comes from the fact that there many different intermediate experiments that will generate the same final updated beliefs. This is summarised in the functional equation, given below, that was derived in Lemma 17

$$u(\mu, p^s) \equiv u\left(u\left(\mu, \mathbf{1} - p_{s'}\right), \left(\tfrac{p_s^\theta}{1-p_{s'}^\theta}\right)_{\theta\in\Theta}\right) \qquad \forall p_{s'}.$$

This says that the update for the signal s, $u(\mu, p^s)$, is also equal to the update for all equivalent intermediate experiments when the signal s' did not occur, then the individual had to do two updates. (They would first revise their beliefs to $u\left(\mu, \mathbf{1} - p_{s'}\right)$ when s' didn't occur. Then, they would do a second update upon observing s.)

One could turn this equation into a family of PDE's and thereby determine u. However, there are techniques to solve this particular functional equation without the assumption of differentiability—continuity is enough. The first step is to turn it into a linear functional equation by taking logarithms. Giving the equation

$$u(\mu, p^s) \equiv u\left(u\left(\mu, \mathbf{1} - p_{s'}\right), p_s - (\mathbf{1} - p_{s'})\right).$$

One simple solution to this functional equation is to add the arguments together: $u(\mu, p) = \mu + p$. Suitably adapted to the fact that μ, p^s, and $u(.)$ are vectors of probabilities and the homogeneity result above, this simple solution gives Bayes' updating. For a more general solution one can first notice that the functional equation tells us about the contours of the function $u(.)$. This is because as $p_{s'}$ varies the arguments on right, that is $u\left(\mu, \mathbf{1} - p_{s'}\right)$ and $p_s - (\mathbf{1} - p_{s'})$ vary in a way that does not change the value of u. It is, then, relatively simple to see that each such contour of the function $u(.)$ is a translation of the other. Thus once the form of one contour has been determined all other contours are just translations of it. Hence choosing an arbitrary function to determine the shape of one contour and another to determine the value taken by each contour is sufficient to determine the entire function. This is how the family of solutions to the functional equation given in the Proposition are generated.

The role of Axiom 3 here is to ensure that we are looking for a continuous solution to the functional equation, and that it holds globally not just locally. If continuity fails the set of solutions to even the simplest functional equation becomes enormous.[14] The role of uniqueness is to ensure a global solution exists, F is a homeomorphism, otherwise there may be many local solutions to the functional equation and only local solutions exist. These local solutions also have the same form as the global solution, see Berg (1993).

Proposition 1. If the updating $\mathcal{U}_n$ satisfies the Axioms 1, 2, 3, and 4, then there exists a homeomorphism $F : \Delta(\Theta) \to \Delta(\Theta)$ such that

$$u(\mu, p_s) = F^{-1}\left(\frac{F_1(\mu) p_s^1}{\sum_{\theta \in \Theta} F_\theta(\mu) p_s^\theta}, \ldots, \frac{F_{|\Theta|}(\mu) p_s^{|\Theta|}}{\sum_{\theta \in \Theta} F_\theta(\mu) p_s^\theta} \right); \tag{8}$$

where $F(\mu) \equiv (F_1(\mu), F_2(\mu), \ldots, F_{|\Theta|}(\mu))$, and

$$\mathcal{U}_n(\mu, \mathcal{E}) = (u(\mu, p_1), \ldots, u(\mu, p_n)) \,.$$

Proof. We begin by reducing the dimension of the variables in the function $u(\mu, p_s)$. Define $w : \Delta^o(\Theta) \to \mathbb{R}_{++}^{|\Theta|-1}$ as follows

$$w(\mu_1, \ldots, \mu_{|\Theta|}) := \left(\frac{\mu_1}{\mu_{|\Theta|}}, \ldots, \frac{\mu_{|\Theta|-1}}{\mu_{|\Theta|}} \right);$$

(where $\mathbb{R}_{++} := \{x \in \mathbb{R} : x > 0\}$). The function w is a bijection, from $\Delta^o(\Theta)$ to $\mathbb{R}_{++}^{|\Theta|-1}$ and it has the inverse

$$w^{-1}(x_1, \ldots, x_{|\Theta|-1}) = \left(\frac{x_1}{1 + \sum_{i=1}^{|\Theta|-1} x_i}, \ldots, \frac{x_{|\Theta|-1}}{1 + \sum_{i=1}^{|\Theta|-1} x_i}, \frac{1}{1 + \sum_{i=1}^{|\Theta|-1} x_i} \right).$$

We will now re-define the variables of the function u. First we define the variable $\phi := w(\mu)$, which requires us to restrict $\mu \in \Delta^o(\Theta)$ and use continuity to define $u(\mu, p_s)$ when μ is on the boundary of this set. We will also define $\pi := w(p_s)$. The fact $u(\mu, p_s)$ is homogeneous degree zero in $p_s \in (0,1)^{|\Theta|}$, by Lemma 17, implies that there is no loss by this transformation. We will also apply the transformation w to the function and define $v(\phi, \pi) \equiv w(u(\mu, p_s))$. Hence we have transformed a function $u : \Delta^o(\Theta) \times \Delta^o(\Theta) \to \Delta^o(\Theta)$ to a function $v : \mathbb{R}_{++}^{|\Theta|-1} \times \mathbb{R}_{++}^{|\Theta|-1} \to \mathbb{R}_{++}^{|\Theta|-1}$.

If we re-write (7) with this new notation we get

$$v(\phi, \pi) \equiv v(v(\phi, \rho), \pi/\rho), \qquad v : \mathbb{R}_{++}^{|\Theta|-1} \times \mathbb{R}_{++}^{|\Theta|-1} \to \mathbb{R}_{++}^{|\Theta|-1};$$

where $\rho := w(\mathbf{1} - p_2)$ and $\pi/\rho := w((\frac{p_1^\theta}{1-p_2^\theta})_{\theta \in \Theta})$.

[14] See Aczél and Dhombres (1989) Chapter 1 for an example of this.

Now we take logarithms to do a final transformation of this function. Let us define $\tilde{\phi} := \ln \phi$, $\tilde{\pi} := \ln \pi$, $\tilde{\rho} = \ln \rho$, and $\tilde{v}(\tilde{\phi}, \tilde{\pi}) \equiv \ln v(\phi, \pi)$. Then, a final re-writing of (7) with this new notation gives

$$\tilde{v}(\tilde{\phi}, \tilde{\pi}) \equiv \tilde{v}(\tilde{v}(\tilde{\phi}, \tilde{\rho}), \tilde{\pi} - \tilde{\rho}), \qquad \tilde{v} : \mathbb{R}^{|\Theta|-1} \times \mathbb{R}^{|\Theta|-1} \to \mathbb{R}^{|\Theta|-1}.$$

If we define $y = \tilde{\pi} - \tilde{\rho}$ and $z = \tilde{\rho}$ this then becomes the functional equation

$$\tilde{v}(\tilde{\phi}, y + z) \equiv \tilde{v}\left(\tilde{v}(\tilde{\phi}, z), y\right), \qquad \forall \phi, y, z \in \mathbb{R}^{|\Theta|-1}.$$

This is called the translation equation and was originally solved in its multivariate form by Aczél and Hosszú (1956). Given the uniqueness property in the regularity assumption, Axiom 3, the results described in Moszner (1995) page 21 apply. Thus any continuous solution to this equation has the property that here exists a continuous bijection $g : \mathbb{R}^{|\Theta|-1} \to \mathbb{R}^{|\Theta|-1}$ such that

$$\tilde{v}(\tilde{\phi}, \tilde{\pi}) = g^{-1}[g(\tilde{\phi}) + \tilde{\pi}]. \tag{9}$$

Now we will reverse the transformations of this problem. First we will remove the logarithms in (9) to get

$$\ln v(\phi, \pi) = g^{-1}[g(\ln \phi) + \ln \pi].$$

Then we will invert the g^{-1} function

$$g \circ \ln v(\phi, \pi) = g(\ln \phi) + \ln \pi.$$

Now we will introduce the function $h(x) = g(\ln x) = (h_1(x), \ldots, h_{|\Theta|-1}(x))$ to simplify this expression.

$$\begin{aligned} h \circ v(\phi, \pi) &= h(\phi) + \ln \pi \\ h \circ v(\phi, \pi) &= \ln e^{h(\phi)} + \ln \pi \\ h \circ v(\phi, \pi) &= \ln \left(e^{h_1(\phi)} \pi_1, \ldots, e^{h_{|\Theta|1}(\phi)} \pi_{|\Theta|-1}\right) \\ e^{h \circ v(\phi, \pi)} &= \left(e^{h_1(\phi)} \pi_1, \ldots, e^{h_{|\Theta|-1}(\phi)} \pi_{|\Theta|-1}\right) \end{aligned}$$

Finally, we define $f(x) \equiv e^{h(x)}$ and this becomes

$$f(v(\phi, \pi)) = \left(f_1(\phi)\pi_1, \ldots, f_{|\Theta|-1}(\phi)\pi_{|\Theta|-1}\right).$$

Now substitute the $w(.)$ transformation to get

$$f \circ w \circ u(\mu, p_s) = \left(f_1(w(\mu)) \frac{p_s^1}{p^{|\Theta|}|\Theta|_s}, \ldots, f_{|\Theta|-1}(w(\mu)) \frac{p_s^{|\Theta|-1}}{p_s^{|\Theta|}} \right). \tag{10}$$

We will now define the function $F : \Delta^o(\Theta) \to \Delta^o(\Theta)$ so that the following diagram commutes, that is, $f \circ w \equiv w \circ F$. This is possible as w is invertible.

$$\begin{array}{ccc} \Delta^o(\Theta) & \xrightarrow{F} & \Delta^o(\Theta) \\ \downarrow w & & \downarrow w \\ \mathbb{R}_{++}^{|\Theta|-1} & \xrightarrow{f} & \mathbb{R}_{++}^{|\Theta|-1} \end{array}$$

F is a bijection and continuous hence it is a homeomorphism. We will extend this definition of F where necessary to the boundary of $\Delta(\Theta)$ using continuity. Using this we can re-write (10) as

$$w \circ F \circ u(\mu, p_s) = \left(w_1(F(\mu)) \frac{p_s^1}{p_s^{|\Theta|}}, \ldots, w_{|\Theta|-1}(F(\mu)) \frac{p_s^{|\Theta|-1}}{p_s^{|\Theta|}} \right),$$
$$= \left(\frac{F_1(\mu) p_s^1}{F_{|\Theta|}(\mu) p_s^{|\Theta|}}, \ldots, \frac{F_{|\Theta|-1}(\mu) p_s^{|\Theta|-1}}{F_{|\Theta|}(\mu) p_s^{|\Theta|}} \right).$$

Now applying w^{-1} to both sides this gives

$$F \circ u\,(\mu, p_s) \equiv \left(\frac{F_1(\mu) p_s^1}{\sum_{\theta \in \Theta} F_\theta(\mu) p_s^\theta}, \ldots, \frac{F_{|\Theta|}(\mu) p_s^{|\Theta|}}{\sum_{\theta \in \Theta} F_\theta(\mu) p_s^\theta} \right).$$

Applying F^{-1} to both sides of this gives equation (8) in the proposition. The other displayed equation in the proposition follows from a substitution of (8) into (4) □

As an example of this updating rule in action, consider again the model of the arrival of busses in Section 1.1. Recall that μ was the probability the individual attached to the good state and that a Bayesian updater would have a posterior $\frac{\alpha^t \mu}{\alpha^t \mu + \beta^t (1-\mu)}$ if there had been t periods without a bus arriving.

An individual who updates divisibly is described by a homeomorphism $F : (\mu, 1-\mu) \mapsto (F_1(\mu), F_2(1-\mu))$ and its inverse $F^{-1} : (\mu, 1-\mu) \mapsto (F_1^{-1}(\mu), F_2^{-1}(1-\mu))$. If such an individual arrived at the bus stop and learnt that there had been t periods without a bus arriving, then they would update their belief about the good state to

$$\mu_t = F_1^{-1} \left(\frac{F_1(\mu) \alpha^t}{F_1(\mu) \alpha^t + F_2(1-\mu) \beta^t} \right).$$

One more period without a bus would then lead the individual to do the further update

$$\mu_{t+1} = F_1^{-1} \left(\frac{F_1(\mu_t) \alpha}{F_1(\mu_t) \alpha + F_2(1-\mu_t) \beta} \right).$$

However, the updating rule for μ_t given here implies that

$$F_1(\mu_t) = \frac{F_1(\mu) \alpha^t}{F_1(\mu) \alpha^t + F_2(1-\mu) \beta^t}, \qquad \text{and} \qquad F_2(1-\mu_t) = 1 - F_1(\mu_t).$$

If these are substituted into the expression for μ_{t+1} we get

$$\mu_{t+1} = F_1^{-1} \left(\frac{F_1(\mu) \alpha^{t+1}}{F_1(\mu) \alpha^{t+1} + F_2(1-\mu) \beta^{t+1}} \right).$$

Thus the beliefs μ_{t+1} of our individual are exactly those of a new arrival at the bus stop.

We now give two general classes of homeomorphisms F that generate particular classes of divisible updating rules.

4.1. Geometric Weighting

The first example of a homeomorphism $F : \Delta(\Theta) \to \Delta(\Theta)$ we consider results in an updating function that has already been used to model overreaction or under-reaction to new information.

$$F(\mu) = \left(\frac{\mu_1^{\alpha_1}}{\sum_\theta \mu_\theta^{\alpha_\theta}}, \dots, \frac{\mu_{|\Theta|}^{\alpha_{|\Theta|}}}{\sum_\theta \mu_\theta^{\alpha_\theta}} \right);$$

An explicit form for F^{-1} only exists when $\alpha = \alpha_\theta$ (for all θ) .

$$F^{-1}(x) = \left(\frac{\mu_1^{1/\alpha}}{\sum_\theta \mu_\theta^{1/\alpha}}, \dots, \frac{\mu_{|\Theta|}^{1/\alpha}}{\sum_\theta \mu_\theta^{1/\alpha}} \right), \qquad \text{for } \alpha = \alpha_\theta, \quad \forall \theta.$$

We can now simply substitute this functional form into (8) to generate a belief updating rule that satisfies our Axioms. In general this gives the updating rule

$$u\left(\mu, (p_s^\theta)_{\theta\in\Theta}\right) \equiv F^{-1}\left(\frac{\mu_1^{\alpha_1} p_s^1}{\sum_{\theta\in\Theta} \mu_\theta^{\alpha_\theta} p_s^\theta}, \dots, \frac{\mu_{|\Theta|}^{\alpha_{|\Theta|}} p_s^{|\Theta|}}{\sum_{\theta\in\Theta} \mu_\theta^{\alpha_\theta} p_s^\theta} \right).$$

When $\alpha = \alpha_\theta$ for all θ

$$u\left(\mu, (p_s^\theta)_{\theta\in\Theta}\right) \equiv \left(\frac{\mu_1 (p_s^1)^{1/\alpha}}{\sum_{\theta\in\Theta} \mu_\theta (p_s^\theta)^{1/\alpha}}, \dots, \frac{\mu_{|\Theta|} (p_s^{|\Theta|})^{1/\alpha}}{\sum_{\theta\in\Theta} \mu_\theta (p_s^\theta)^{1/\alpha}} \right).$$

From this expression it is clear that this updating rule exaggerates or alters the probabilities that enter into the normal Bayesian formula. This can be interpreted as overreaction or under-reaction to new information. But, in Section 5.2, we will also show that this generates a bias in the updating.

When $\alpha \neq \alpha_\theta$, there is no explicit solution for the full updated distribution. However, it is possible to proceed in a slightly different way to understand how the relative probabilities are updated. First, write

$$F \circ u\left(\mu, (p_s^\theta)_{\theta\in\Theta}\right) \equiv \left(\frac{\mu_1^{\alpha_1} p_s^1}{\sum_{\theta\in\Theta} \mu_\theta^{\alpha_\theta} p_s^\theta}, \dots, \frac{\mu_{|\Theta|}^{\alpha_{|\Theta|}} p_s^{|\Theta|}}{\sum_{\theta\in\Theta} \mu_\theta^{\alpha_\theta} p_s^\theta} \right)$$

and then define $(\mu'_1, \dots, \mu'_{|\Theta|}) = u\left(\mu, (p_s^\theta)_{\theta\in\Theta}\right)$ to get

$$\left(\frac{(\mu'_1)^{\alpha_1}}{\sum_\theta (\mu'_\theta)^{\alpha_\theta}}, \dots, \frac{(\mu'_{|\Theta|})^{\alpha_{|\Theta|}}}{\sum_\theta (\mu'_\theta)^{\alpha_\theta}} \right) \equiv \left(\frac{\mu_1^{\alpha_1} p_s^1}{\sum_{\theta\in\Theta} \mu_\theta^{\alpha_\theta} p_s^\theta}, \dots, \frac{\mu_{|\Theta|}^{\alpha_{|\Theta|}} p_s^{|\Theta|}}{\sum_{\theta\in\Theta} \mu_\theta^{\alpha_\theta} p_s^\theta} \right).$$

Then dividing the θ'' entry in this vector by the θ' entry we get

$$\frac{u_{\theta''}(\mu, (p_s^\theta)_{\theta\in\Theta})^{\alpha_{\theta''}}}{u_{\theta'}(\mu, (p_s^\theta)_{\theta\in\Theta})^{\alpha_{\theta'}}} = \frac{\mu_{\theta''}^{\alpha_{\theta''}}}{\mu_{\theta'}^{\alpha_{\theta'}}} \frac{p_s^{\theta''}}{p_s^{\theta'}}; \qquad \frac{u_{\theta''}(\mu, (p_s^\theta)_{\theta\in\Theta})}{u_{\theta'}(\mu, (p_s^\theta)_{\theta\in\Theta})} = \underbrace{\frac{\mu_{\theta''}}{\mu_{\theta'}} \left(\frac{p_s^{\theta''}}{p_s^{\theta'}} \right)^{1/\alpha}}_{\text{if } \alpha=\alpha_{\theta''}=\alpha_{\theta'}}.$$

In the case where the parameters α_θ differ it is still possible to give an explicit expression for the relative size of the updated belief in θ' and θ''

This updating rule is generalises the overconfidence/under-confidence protocol described in Angrisani, Guarino, Jehiel, and Kitagawa (2017) and Bohren and Hauser (2017), for example. If $1/\alpha_\theta > 1$ then the individual is overreacts to new information on parameter value θ and is overconfident by placing too much weight on their observations. This is achieved by exaggerating the differences in the signals and is more clearly seen in the ratio of the updated beliefs given above. Conversely, if $1/\alpha_\theta < 1$ then the individual is under-reacts to new information about their learning on parameter value θ—they place too much weight on their prior and do not adjust their beliefs as much as a Bayesian would. In these two papers, the under and over-reaction is uniform across all parameters. But, the functional form permits the agent to be over-react for some parameters and under-react for others. Thus there is the possibility of selective over and under-reaction where the agent more readily changes beliefs about some parameters but not about others.

4.2. *Exponential Weighting and Other Homeomorphisms*

For all $\theta \in \Theta$, let $f_\theta : [0,1] \to \mathbb{R}_+$ be a strictly increasing and continuous function that satisfies $f_\theta(0) = 0$. Then let us define F, as below

$$F(\mu) = \left(\frac{f_1(\mu_1)}{\sum_\theta f_\theta(\mu_\theta)}, \ldots, \frac{f_{|\Theta|}(\mu_{|\Theta|})}{\sum_\theta f_\theta(\mu_\theta)} \right).$$

Without an explicit form for F^{-1} we can use the relation

$$F \circ u \left(\mu, (p_s^\theta)_{\theta \in \Theta} \right) \equiv \left(\frac{f_1(\mu_1) p_s^1}{\sum_{\theta \in \Theta} f_\theta(\mu_\theta) p_s^\theta}, \ldots, \frac{f_{|\Theta|}(\mu_{|\Theta|}) p_s^{|\Theta|}}{\sum_{\theta \in \Theta} f_\theta(\mu_\theta) p_s^\theta} \right).$$

Again if we write the updated beliefs as $u\left(\mu, (p_s^\theta)_{\theta \in \Theta}\right) = (\mu'_1, \ldots, \mu'_{|\Theta|})$, then the ratio of any two entries gives

$$\frac{f_\theta(\mu'_\theta)}{f_{\theta'}(\mu'_{\theta'})} = \frac{f_\theta(\mu_\theta) p_s^\theta}{f_{\theta'}(\mu_{\theta'}) p_s^{\theta'}}. \tag{11}$$

As an example of such a function we could choose $f_\theta(\mu) = e^{-\beta_\theta/\mu}$, which results in a transformation of beliefs that is similar to a multinomial logit.

$$F(\mu) = \left(\frac{e^{-\beta_1/\mu_1}}{\sum_\theta e^{-\beta_\theta/\mu_\theta}}, \ldots, \frac{e^{-\beta_{|\Theta|}/\mu_{|\Theta|}}}{\sum_\theta e^{-\beta_\theta/\mu_\theta}} \right) \qquad \text{for } \beta_\theta > 0 \ \forall \theta.$$

This function F does not nest Bayesian updating—there are no values of the parameters β for which this function is the identity. Suppose that $\beta_\theta = \beta$ for all θ. Then F does map points closer to $\bar{\mu}$ when β is small and move points away from $\bar{\mu}$ when β is large. As $\beta \to 0$, so $F(\mu)$ converges to $\bar{\mu}$ for all interior μ. And as $\beta \to \infty$, so $F(\mu)$ converges to a distribution that puts all weight on the largest elements of μ. Thus the extremes of this function are similar to those in the previous example. And, using the intuitions from our previous example, we would expect large values of β to be associated with under-reaction and small values of β to be associated with over-reaction to new information.

An explicit form for the inverse function F^{-1} is not given here. However, by taking the ratio of any two entries (say $\hat\theta$ and $\tilde\theta$) in these vectors we get

$$\frac{e^{-\beta_{\hat\theta}/\mu'_{\hat\theta}}}{e^{-\beta_{\tilde\theta}/\mu'_{\tilde\theta}}} = \frac{e^{-\beta_{\hat\theta}/\mu_{\hat\theta}}p_s^{\hat\theta}}{e^{-\beta_{\tilde\theta}/\mu_{\tilde\theta}}p_s^{\tilde\theta}}.$$

Alternatively

$$\frac{\beta_{\tilde\theta}}{\mu'_{\tilde\theta}} - \frac{\beta_{\hat\theta}}{\mu'_{\hat\theta}} = \frac{\beta_{\tilde\theta}}{\mu_{\tilde\theta}} - \frac{\beta_{\hat\theta}}{\mu_{\hat\theta}} + \ln\frac{p_s^{\hat\theta}}{p_s^{\tilde\theta}}.$$

Thus the updating of beliefs results in a linear shift in the inverse probabilities of each parameter value. If $\beta_{\hat\theta} = \beta_{\tilde\theta} = \beta$ it is easy to see that large values of β decrease the dependence of this shift on on the ratio of the probabilities. Thus there is the conjectured under-reaction in this case.

4.3. Trigonometric Updating

Consider the function $F : (\mu, 1-\mu) \mapsto (\sin^2\frac{\pi}{2}\mu, \sin^2\frac{\pi}{2}(1-\mu))$. This is a homeomorphism from $\Delta(\{\theta,\theta'\})$ into itself as $\sin\frac{\pi}{2}(1-\mu) = \cos\frac{\pi}{2}\mu$. Thus F characterises a divisible updating rule for dichotomous experiments. If this functional form is substituted into (11) we get The updating function is

$$\frac{\sin^2\frac{\pi}{2}\mu'_\theta}{\sin^2\frac{\pi}{2}(1-\mu'_\theta)} = \frac{p_s^\theta}{p_{s'}^\theta}\frac{\sin^2\frac{\pi}{2}\mu_\theta}{\sin^2\frac{\pi}{2}(1-\mu_\theta)}$$

Then applying the identity $\sin(\frac{\pi}{2}-x) = \cos x$ this becomes

$$\tan\left(\frac{\pi}{2}\mu'_\theta\right) = \left(\frac{p_s^\theta}{p_{s'}^\theta}\right)^{1/2}\tan\left(\frac{\pi}{2}\mu_\theta\right).$$

Finally this gives

$$\mu'_\theta = \frac{2}{\pi}\arctan\left(\frac{\sqrt{p_s^\theta}}{\sqrt{p_{s'}^\theta}}\tan\left(\frac{\pi}{2}\mu_\theta\right)\right).$$

5. Properties of Divisible Updating

In this section we describe some of the general properties of the updating rules that satisfy the Axioms 1, 2, 3, and 4. In particular we will show that these updating rules obey consistency and describe conditions on the function F that ensure updating is biased in particular ways.

We are going to need to impose the condition that the updating respects certainty. That is, if the individual has a prior that attaches probability one to the parameter θ and observes signals that only occur with positive probability under this prior, then they do not revise their beliefs. We let $e_\theta \in \Delta(\Theta)$ denote the prior that attaches probability one to the parameter θ, that is, it is a vector with unity in the θ entry and zeros elsewhere.

$$\mathcal{U}_n(e_\theta, \mathcal{E}) = (e_\theta, \ldots, e_\theta), \qquad \text{for all } \mathcal{E} \in \Delta^o(S)^{|\Theta|}. \tag{12}$$

5.1. *The Consistency of Divisible Updating*

In this section we consider the limiting properties of the updating rule (8) as increasing amounts of information are collected. That is, we will consider the limit of the updated beliefs as the experiment $\mathcal{E}$ is repeatedly sampled for a given value of the parameter. The main result is that the updating converges almost surely. And, if the experiment is informative, then the updated belief converges with probability one to certainty that the parameter is θ. This property of updating is usually termed *consistency* when it holds in the Bayesian case, see Diaconis and Freedman (1986) for example. The consistency of divisible updating contrasts with other examples of non-Bayesian updating in the literature that do not satisfy consistency.[15]

In a stationary environment there are two properties that might be desirable in a model of updating: (1) the individual believes they will learn (they attach probability one to their updated beliefs converging), (2) they do actually learn (their updated beliefs do converge with probability one). In general there are updating procedures where neither, all, or one of these properties hold. The result below shows that both of the above properties are satisfied by divisible learning when the model is correctly specified.

We will begin with a description of the model where the individual repeatedly samples from the same experiment for a given parameter value. Fix a parameter value θ and an experiment $\mathcal{E} = (p^\theta)_{\theta\in\Theta}$. Let the stochastic process $\{s^t\}_{t=0}^\infty \in S^\infty$ be independently sampled from the distribution $p^\theta \in \Delta^o(S)$. We will use $\mathbb{P}^\theta$ to denote the probability measure on S^∞ induced by this process. We will also inductively define the stochastic process $\{\mu^t\}_{t=0}^\infty \in \Delta(\Theta)^\infty$ so that μ^{t+1} is the updated value of μ^t when the signal s^t is observed. The formal definition is: $\mu^0 \in \Delta(\Theta)$ and

$$\mu^{t+1} := u\left(\mu^t, p_{s^t}\right), \qquad t = 0, 1, \ldots \tag{13}$$

Proposition 2 shows that for all θ the process $\{\mu^t\}_{t=0}^\infty$ converges $\mathbb{P}^\theta$ almost surely provided the updating satisfies Axioms 1, 2, 3, 4. Furthermore, it shows that if the updating satisfies condition (12) and the signals are informative, then μ^t converges to e_θ, $\mathbb{P}^\theta$ almost surely. That is the updating satisfies consistency.

[15]See Rabin and Schrag (1999) and Epstein, Noor, and Sandroni (2010) for examples of inconsistent updating. In Lehrer and Teper (2015) the notion of consistency is used as an axiom to characterise Bayesian updating.

Proposition 2. If the updating $\mathcal{U}_n$ satisfies the Axioms 1, 2, 3, 4, then for all θ there exists $\mu^\infty \in \Delta(\Theta)$ such that $\mu^t \to \mu^\infty$, $\mathbb{P}^\theta$ almost surely. If (a) $p^\theta \neq p^{\theta'}$ for all $\theta' \neq \theta$, (b) $\mu^0 \in \Delta^o(\Theta)$, and (c) (12) holds, then $\mu^\infty = e_\theta$ with $\mathbb{P}^\theta$ probability one.

The proof of this result is given in the appendix. The proof is quite trivial; it just applies the usual proof of the consistency of Bayesian updating (see for example DeGroot (1970)) to the shadow-belief revision process. It then uses the property that belief revision respects certainty (12) to ensure that when the shadow beliefs approach certainty they are mapped back (by a continuous inverse function) to beliefs that also approach certainty.

5.2. *Bias in Divisible Updating*

In this section we describe the classes of divisible updating rules that exhibit two kinds of bias. The first type of bias (which we term "local consistency") is that if θ is the true parameter, then the expected value of the updated belief in θ is greater than the original belief:

$$\mu_\theta \leq E^\theta(u_\theta(\mu, p_s)) \,.$$

E^θ is an expectation take relative to the objective distribution of the signals, so the condition says that an observer expects the individual with initial beliefs μ^θ to have an increased belief in θ when they observe the outcome of the experiment $\mathcal{E}$ and θ is true. This condition holds globally for a Bayesian updater and is the well-known conditional submartingale property for Bayesian posteriors.

The second bias, which we term "local inconsistency", is the reverse of the above

$$\mu_\theta \geq E^\theta(u_\theta(\mu, p_s)) \,.$$

When this inequality holds an observer expects the individual to have an updated belief in θ that is less than their original. This kind of bias does not arise because individuals are ignoring or misinterpreting their signals. It arises because when θ is true the individual is slow to move their belief in θ upwards in response to positive evidence but quick to move beliefs down when evidence in favour of an alternative θ' is observed. These two effects give, on average, a downward movement of beliefs. Thus, the bias could be interpreted as a reluctance to move to extreme beliefs or a sceptical attitude to extreme evidence. With divisible updating this sceptical attitude cannot hold at all priors: it must be a local not a global property. This is because (by the results in the previous section) beliefs must ultimately converge to the truth for all divisible updating processes.

For tractability, the result on biases given here will apply only to experiments that have two possible parameters $\Theta = \{\theta, \theta'\}$: such experiments are also called dichotomies. In dichotomies the homeomorphism $F : (\mu_\theta, \mu_{\theta'}) \mapsto (\mu'_\theta, \mu'_{\theta'})$ can be described by its effect on its first element:

$$F(\mu) \equiv (f(\mu_\theta), 1 - f(\mu_\theta)).$$

Where $f : [0, 1] \to [0, 1]$ is strictly increasing and continuous (as F is a homeomorphism). The belief updating of θ conditional on the signal s can then be written explicitly as

$$u_\theta(\mu, p_s) = f^{-1} \circ \frac{f(\mu_\theta)p_s^\theta}{f(\mu_\theta)p_s^\theta + (1 - f(\mu_\theta))p_s^{\theta'}}. \tag{14}$$

In the result below we show that one of the two biases occur if the function $1/f(.)$ is locally either convex or concave. Hence, we need a notion of a neighborhood of the original belief $\mu_\theta \in (0, 1)$. Define the interval $R_f(\mu_\theta) \subset (0, 1)$ so that $\mu_\theta \in R_f(\mu_\theta)$ and $u_\theta(\mu, p_s) \in R_f(\mu_\theta)$ for all $s \in S$. This ensures that $R_f(\mu_\theta)$ includes all possible values of the updated beliefs in θ when the prior is μ_θ. Equipped with this definition we can provide sufficient conditions for the biases described above. The proof of this proposition is given in the appendix.

Proposition 3. Suppose the divisible updating in a dichotomy is described by (14), where $f(.)$ is continuous and strictly increasing. Then,

(i) If $1/f(.)$ is convex on $R_f(\mu_\theta)$, then $\mu_\theta \leq E^\theta(u_\theta(\mu, p_s))$.

(ii) If $1/f(.)$ is concave on $R_f(\mu_\theta)$, then $\mu_\theta \geq E^\theta(u_\theta(\mu, p_s))$.

If $f(0) = 0$ then $1/f(.)$ is not concave on any interval of the form $(0, x)$.

For Bayesian updating $f(\mu)$ is the identity ($1/f(\mu) = \mu^{-1}$ is convex). In this case Proposition 3 is the usual result that Bayesian updating is a conditional submartingale and on average are revised upwards.

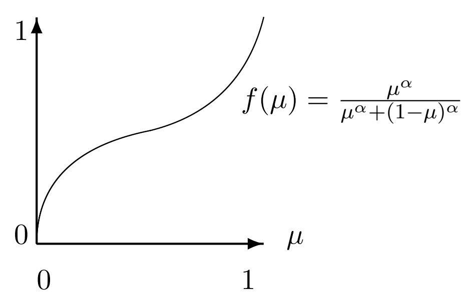

FIGURE 3. $\frac{\mu^\alpha}{\mu^\alpha + (1-\mu)^\alpha}$, for $\alpha < 1$.

For an example, consider the divisible updating that is given in Section 4.1 where F has a geometric weighting form (and we give each of the two parameters equal weight $\alpha_\theta = \alpha_{\theta'} = \alpha$). In dichotomies this gives

$$f(\mu) = \frac{\mu^\alpha}{\mu^\alpha + (1 - \mu)^\alpha}, \qquad u_\theta(\mu, p_s) = \frac{\mu(p_s^\theta)^{1/\alpha}}{\mu(p_s^\theta)^{1/\alpha} + (1 - \mu)(p_s^{\theta'})^{1/\alpha}}.$$

Calculus shows that $1/f(\mu)$ is convex when $\alpha \geq 1$, so,this mode of updating is locally, and globally, consistent. However, when $\alpha < 1$ there exists an interval of values $\mu \in (\frac{1}{2}(1+\alpha), 1]$ where $1/f(\mu)$ is concave, so for high values of μ the updating is locally inconsistent. As the individual gets closer to certain that θ is the true parameter, an outside observer would expect that the individual's belief in θ would decrease on average. The convergence to certainty will be slower as a result of this downward drift. Inspection of the expression for $u_\theta(\mu, p_s)$ above shows that the updated beliefs become more sensitive to the signal probabilities as α declines, thus it appears that this downward bias in the updating is related to the increased sensitivity of this updating to the current signals. This overreaction to information is something that we will investigate in the next section. An alternative interpretation of this effect is to notice that when $\alpha < 1$ the function $f(.)$ maps beliefs to more central positions away from the extremities of the interval $[0, 1]$ and it is this the feature that generates the local inconsistency at high beliefs.

5.3. *Sufficient Conditions for Under and Overreaction to Information*

In this section sufficient conditions for the divisible updating rule to overreact or under-react to new information are given. Again, for tractability, we will restrict attention to dichotomies and consider updating that is characterised by a strictly increasing function $f : [0, 1] \to [0, 1]$ as in (14). We will provide sufficient conditions on the function $f(.)$ for the updating it describes to be more or less variable than a Bayesian's and then end with some examples.

Overreaction has many meanings in the literature on updating; here it is defined to mean that the updated beliefs have a greater variance than those of Bayesian updater who faces the same experiment. The log-likelihood ratio of a Bayesian updater follows a homogeneous random walk. Thus a prior-independent measure of the variability of Bayesian updating is the variance of the log-likelihood ratio. This can be seen from the simple calculation

$$\mathrm{Var}\left[\ln \frac{\mu'_\theta}{1-\mu'_\theta}\right] = \mathrm{Var}\left[\ln \frac{\mu_\theta}{1-\mu_\theta} + \ln \frac{p_s^\theta}{p_s^{\theta'}}\right] = \mathrm{Var}\left[\ln \frac{p^\theta}{p^{\theta'}}\right].$$

(Here μ'_θ denotes the Bayesian update and the variance is taken unconditionally of the parameter value.) Thus $\mathrm{Var}\left[\ln \frac{p^\theta}{p^{\theta'}}\right]$ will be our benchmark and we will say the updating $u_\theta(\mu, p_s)$ in (14) exhibits overreaction if

$$\mathrm{Var}\left[\ln \frac{\mathrm{u}_\theta(\mu, \mathrm{p_s})}{1-\mathrm{u}_\theta(\mu, \mathrm{p_s})}\right] > \mathrm{Var}\left[\ln \frac{\mathrm{p}^\theta}{\mathrm{p}^{\theta'}}\right], \qquad \forall \mu.$$

Similarly, we will say that the updating exhibits under-reaction if

$$\mathrm{Var}\left[\ln \frac{\mathrm{u}_\theta(\mu, \mathrm{p_s})}{1-\mathrm{u}_\theta(\mu, \mathrm{p_s})}\right] < \mathrm{Var}\left[\ln \frac{\mathrm{p}^\theta}{\mathrm{p}^{\theta'}}\right], \qquad \forall \mu.$$

There is a simple intuition for the result below. If the inverse homeomorphism F^{-1}, described in Figure 2, expands space then when the shadow posterior and prior are mapped back to the belief space they are even further apart. The response to the signals has become more exaggerated and overreaction is present. Similarly, if the function F^{-1} contracts space, then the learning that occurred in the shadow Bayesian world gets reduced when it is mapped back to the belief space by F^{-1}. As a result the Bayesian learning in the shadow space is understated and there is under-reaction to new information. In the one-dimensional case this means that the slope of the function $f(.)$ will play a role in characterising under- or overreaction.

We now are able to state a result on when divisible updating exhibits over and under-reaction in the two-parameter case.

Proposition 4. Suppose that the divisible updating in a dichotomy, $u_\theta(\mu, p_s)$, is described by the function $f(.)$, as in (14), and that f is continuously differentiable.
If $f'(\mu) > \frac{f(\mu)(1-f(\mu))}{\mu(1-\mu)}$ for all $\mu \in (0,1)$, the updating exhibits under-reaction.
If $f'(\mu) < \frac{f(\mu)(1-f(\mu))}{\mu(1-\mu)}$ for all $\mu \in (0,1)$, the updating exhibits overreaction.

Two examples of divisible updating that we could investigate using this result are geometric weighting and exponential weighting. Their associated f functions for dichotomies (respectively denoted f_g, f_e) are defined below.

$$f_g(\mu) := \frac{\mu^\alpha}{\mu^\alpha + (1-\mu)^\alpha}, \qquad \text{and} \qquad f_e(\mu) := \frac{e^{-\beta/\mu}}{e^{-\beta/\mu} + e^{-\beta/(1-\mu)}}.$$

Some calculus shows that

$$f'_g(\mu) = \alpha \frac{f_g(\mu)(1-f_g(\mu))}{\mu(1-\mu)} \qquad \text{and} \qquad f'_e(\mu) = \beta f_e(\mu)(1-f_e(\mu)) \left(\frac{1}{\mu^2} + \frac{1}{(1-\mu)^2} \right).$$

Thus a simple application of Proposition 4 says that geometric updating exhibits under-reaction if $\alpha > 1$ and over reaction if $\alpha < 1$. Whereas exponential updating exhibits under reaction if $\beta > 1/2$ and can never satisfy the global conditions for overreaction.

6. Characterisation of Bayesian Updating

In this section we show that one additional axiom—that the updating is unbiased, or a martingale, or satisfies Bayes' plausibility—is sufficient for full Bayesian updating.

One might view of Proposition 1 as saying that the axioms we have provided so far are "almost" enough for Bayesian updating. Thus it seems likely that a small additional requirement will give characterise Bayesian updating. We do not claim that the martingale property is minimal in this sense. There may well be weaker restrictions on the updating process that when added to Axioms 1, 2, 3, and 4 restrict the updating to Bayesianism. However, the unbiased nature of belief revision is such a fundamental property that it does seem important to consider it directly. Furthermore the non-Bayesian updating of Epstein, Noor, and Sandroni (2010) is a martingale. Hence, it provides a useful example of a belief revision process that satisfies Axioms 1, 2, 3 and the martingale condition but is not Bayesian.[16]

[16]This revision process only satisfies Axiom 3 when there is over-reaction. In the case of underreaction extreme updated beliefs are not possible.

Another route to take in this section would be to try to find an axiom that rules out features of updating that are non-Bayesian but are nevertheless consistent with Proposition 1. For example, to try to find an axiom that does not admit over-reaction or under-reaction to new information. This is not what the martingale axiom does. Indeed the model of Epstein, Noor, and Sandroni (2010) does permit over-reaction or under-reaction and is also a martingale. Thus it appears that it is the interaction of the martingale property and the divisibility property that jointly act to give Bayesian updating.

The Axiom below considers the profile of updated belief distributions: this is the distribution $\mathcal{U}_n^s(\mu,(p)_{\theta\in\Theta}) \in \Delta(\Theta)$ for each signal s. Then it averages these distributions with weights that are the ex-ante probabilities of the signals: $\sum_{\theta\in\Theta}\mu_\theta p_s^\theta$. This average is the decision maker's predicted distribution of updated beliefs and the axiom requires that this predicted distribution equals the original beliefs.

Axiom 5 (Unbiased)**.** For any $\mu > 0$, $n > 1$, and $\mathcal{E} = (p^\theta)_{\theta\in\Theta} \in \Delta^o(S)^{|\Theta|}$ the updating function $\mathcal{U}_n(\mu,\mathcal{E})$ satisfies

$$\mu \equiv \sum_{s\in S}\left(\sum_{\theta\in\Theta}\mu_\theta p_s^\theta\right)\mathcal{U}_n^s(\mu,\mathcal{E}).$$

We will now prove the following result

Proposition 5. If the updating $\mathcal{U}_n$ satisfies the Axioms 1, 2, 3, 4, and 5, then it satisfies Bayes rule, that is, $\mathcal{U}_n(\mu,\mathcal{E}) \equiv \big(u(\mu,p_1),\dots,u(\mu,p_{|\Theta|})\big)$ where

$$u(\mu,p_s) \equiv \left(\frac{\mu p_s^1}{\sum_{\theta\in\Theta}\mu p_s^\theta},\dots,\frac{\mu p_s^{|\Theta|}}{\sum_{\theta\in\Theta}\mu p_s^\theta}\right). \tag{15}$$

Proof. It is sufficient to prove that the function $F(.)$ in Proposition 1 is the identity. Without loss, we will consider the case where $n = 2$ and there is a binary experiment with signals $S = \{1,2\}$. When the parameter is θ we will write the probabilities of the two signals as $(p^\theta, 1-p^\theta)$.

In this case the unbiased condition, Axiom 5, is equivalent to:

$$\mu \equiv \left(\sum_\theta \mu_\theta p^\theta\right)F^{-1}(u_1^1,\dots,u_{|\Theta|}^1) + \left(\sum_\theta \mu_\theta(1-p^\theta)\right)F^{-1}\left(u_1^2,\dots,u_{|\Theta|}^2\right), \tag{16}$$

$$\text{where}\qquad u_{\theta'}^1 = \frac{F_{\theta'}(\mu)p^{\theta'}}{\sum_\theta F_\theta(\mu)p^\theta},\qquad u_{\theta'}^2 = \frac{F_{\theta'}(\mu)(1-p^{\theta'})}{\sum_\theta F_\theta(\mu)(1-p^\theta)},\qquad \forall\theta'\in\Theta.$$

This holds for all values of $\mu \in \Delta^o(\Theta)$, all $p^\theta \in [0,1]$, because $F(.)$ is continuous bounded and can be defined on the boundary of $(0,1)$.

We will begin by showing that F is the identity at the extreme points of the set $\Delta(\Theta)$. Letting $p^{\theta'} \to 0$ in (16) for all $\theta' \neq \theta$ and $p^\theta \to 1$ gives the identity

$$\mu \equiv \mu_\theta F^{-1}(e_\theta) + (1-\mu_\theta)F^{-1}(y_\theta). \tag{17}$$

where: e_θ is a vector with 1 in the θ element and zeros elsewhere and y_θ is a vector with zero in the θ element and $F_{\theta'}(\mu)/(1-F_\theta(\mu))$ for $\theta' \neq \theta$. The θth element of the vector identity (17) is

$$\frac{\mu_\theta}{1-\mu_\theta}(1-F_\theta^{-1}(e_\theta)) \equiv F_\theta^{-1}(y_\theta).$$

If $1 \neq F_\theta^{-1}(e_\theta)$, the left of this expression is arbitrarily large as $\mu_\theta \to 1$, but the right takes values only in $[0,1]$. Thus, this identity holds for all $\mu_\theta \in (0,1)$ only if: $1 = F_\theta^{-1}(e_\theta)$ and $0 = F_\theta^{-1}(y_\theta)$. As $F^{-1}(\mu) \in \Delta(\Theta)$, the first of these implies that $e_\theta = F^{-1}(e_\theta)$ and F is the identity on its extreme points. The second condition, $0 = F_\theta^{-1}(y_\theta)$, implies that F maps zero probabilities to zero probabilities.

We now derive a functional equation (19) which will allow us to establish the Proposition. Letting $p^{\theta'} \to 0$ for all $\theta' \neq \theta$ in (16) and imposing the continuity of F now implies

$$\mu \equiv \mu_\theta p^\theta F^{-1}(e_\theta) + (1-\mu_\theta p^\theta)F^{-1}\left(\frac{F(\mu) - p^\theta F_\theta(\mu)e_\theta}{1-p^\theta F_\theta(\mu)}\right).$$

We have shown that $F^{-1}(e_\theta) = e_\theta$, so this rearranges to give

$$F\left(\frac{\mu - p^\theta \mu_\theta e_\theta}{1-\mu_\theta p^\theta}\right) \equiv \frac{F(\mu) - p^\theta F_\theta(\mu)e_\theta}{1-p^\theta F_\theta(\mu)}. \tag{18}$$

We will reduce the dimension of the function F by dropping one element (μ_θ) of the vector μ and considering it as a function that maps $\mu_{-\theta}$ to $\mu'_{-\theta}$ allowing the remaining probability to be determined by the adding up requirement. Define $\mu_{-\theta} \in \mathcal{S} := \{x \in \mathbb{R}_+^{|\Theta|-1} : \mathbf{1}^T x \leq 1\}$. Now define $G_\theta : \mathcal{S} \to \mathcal{S}$, so that $G_\theta(\theta_{-\theta})$ describes how F maps the vector $\mu_{-\theta}$ into $\mathcal{S}$, as follows

$$F(\mu) \equiv (G_\theta(\mu_{-\theta}), 1 - \mathbf{1}^T G_\theta(\mu_{-\theta})).$$

(As F is a homeomorphism, so too is G.) Dropping the θth row in the vector equation (18)and re-writing what remains using the G_θ notation then gives

$$G_\theta\left(\frac{\mu_{-\theta}}{1-\mu_\theta p^\theta}\right) \equiv \frac{G_\theta(\mu_{-\theta})}{1-p^\theta F_\theta(\mu)}.$$

Finally, defining $\lambda := (1-\mu_\theta p^\theta)^{-1} \in [1, (1-\mu_\theta)^{-1}]$ into the above then gives

$$G_\theta(\lambda \mu_{-\theta}) \equiv \frac{\lambda}{\lambda - (1-\lambda)\frac{F_\theta(\mu)}{\mu_\theta}} G_\theta(\mu_{-\theta}), \qquad \forall\ \lambda \in [1, (1-\mu_\theta)^{-1}]. \tag{19}$$

The function G_θ satisfies a property that is similar to homogeneity (full homogeneity holds when $F_\theta(\mu) = \mu_\theta$).

For any $\mu \in \Delta^o(\Theta)$ consider the sequence $\{\mu^n\}_{n=1}^\infty$ where

$$\mu^n := \left(\frac{\mu_{-\theta}}{n}, \frac{n-1+\mu_\theta}{n}\right).$$

Along this sequence we have that $\mu^n_\theta \to 1$ and $F_\theta(\mu^n) \to 1$ by the previously established result ($F(e_\theta) = e_\theta$) and the continuity of F. Writing the relation (19) for μ_n gives

$$G_\theta\left(\frac{\lambda'}{n}\mu_{-\theta}\right) \equiv \frac{\lambda'}{\lambda' - (1-\lambda')\frac{F_\theta(\mu^n)}{\mu^n_\theta}} G_\theta\left(\frac{1}{n}\mu_{-\theta}\right), \qquad \forall\, 1 \le \lambda' \le \frac{n}{1-\mu_\theta}.$$

If we make the choices $\lambda' = n$ and $\lambda' = n\lambda$, then we can use the left of this expression to substitute for the terms $G_\theta(\lambda\mu_{-\theta})$ and $G_\theta(\mu_{-\theta})$ in (19). This then gives

$$\frac{n}{n-(1-n)\frac{F_\theta(\mu^n)}{\mu^n_\theta}} G_\theta\left(\frac{1}{n}\mu_{-\theta}\right) \equiv \frac{\lambda}{\lambda-(1-\lambda)\frac{F_\theta(\mu)}{\mu_\theta}} \frac{n\lambda}{n\lambda-(1-n\lambda)\frac{F_\theta(\mu^n)}{\mu^n_\theta}} G_\theta\left(\frac{1}{n}\mu_{-\theta}\right).$$

Eliminating G_θ from the above and re-arranging gives

$$1 \equiv \frac{\lambda}{\lambda-(1-\lambda)\frac{F_\theta(\mu)}{\mu_\theta}} \frac{\lambda - \lambda(\frac{1}{n}-1)\frac{F_\theta(\mu^n)}{\mu^n_\theta}}{\lambda-(\frac{1}{n}-\lambda)\frac{F_\theta(\mu^n)}{\mu^n_\theta}}, \qquad \forall n.$$

We have shown that $\mu^n_\theta \to 1$ and $F_\theta(\mu^n) \to 1$ as $n \to \infty$, so this implies that

$$1 = \frac{\lambda}{\lambda-(1-\lambda)\frac{F_\theta(\mu)}{\mu_\theta}}, \qquad \forall\, \lambda \in [1, (1-\mu_\theta)^{-1}].$$

This implies $F_\theta(\mu) = \mu_\theta$. This condition holds for all θ and all $\mu \in \Delta^o(\Theta)$, so $F(\mu) = \mu$ for all $\mu \in \Delta^o(\Theta)$. As F is continuous on $\Delta(\Theta)$ it follows that F is the identity. Substitution of $F(\mu) = \mu$ into (8) establishes the claim in this proposition. □

7. Unbiased Updating

In this section the set of updating rules that satisfy the unbiasedness property for the experiment $\mathcal{E}$ characterised. It is shown that any updating rule in this class can be interpreted as application of Bayes rule to an alternative experiment $\mathcal{E}'$ that has the same unconditional signal probabilities as $\mathcal{E}$ but is otherwise unrestricted.

Let us first re-describe the model of updating using matrix notation. We begin by describing the parameter-dependent signal probabilities as a matrix. Let P be the non-negative $|S| \times |\Theta|$ matrix with columns $p^\theta \in \Delta(S)$; the signal distributions for the parameters θ. If we treat the prior $\mu \in \Delta(\Theta)$ as a column vector, then the product $P\mu \in \Delta(S)$ gives a vector that is the unconditional probabilities of signals for the experiment $\mathcal{E} = P$. The updating function which was has the general form $(\mu^1, \dots, \mu^n) = \mathcal{U}_n(\mu, (p^\theta)_{\theta\in\Theta})$ will now be written as $U = \mathcal{U}(\mu, P)$, where U is a non-negative $|\Theta| \times |S|$ matrix U. The columns of this matrix are $\mu^s \in \Delta(\Theta)$, the updated beliefs after the signal s. The unbiased property for the updating function can now be written as:

$$\mu \equiv \mathcal{U}(\mu, P)P\mu; \qquad \text{where} \qquad \mathbb{1}^T P = \mathbb{1}^T, \;\; \mathbb{1}^T\mathcal{U}(\mu, P) = \mathbb{1}^T, \;\; \mathbb{1}^T\mu = 1.$$

Any function $\mathcal{U}(\mu, P)$ that satisfies these for all experiments (μ, P) describes an updating function with the martingale property. This motivates to the following (equivalent) definition of an unbiased updating function.

Definition 1. *The updating function* $\mathcal{U} : \Delta(\Theta) \times \Delta(S)^{|\Theta|} \to \Delta(\Theta)^n$*, is unbiased if*

$$\mu = \mathcal{U}(\mu, P)P\mu; \tag{20}$$

for all $\mu \in \Delta(\Theta)$ *and all* $P \in \Delta(\Theta)^n$*.*

By way of a benchmark let us describe standard Bayesian updating using this notation. Let $D_{(x)}$ denote the $n \times n$ matrix with the vector $x \in \mathbb{R}^n$ on its diagonal and zeros elsewhere. Given the experiment (μ, P), the $|\Theta| \times |S|$ matrix $D_{(\mu)}P^T$ is therefore the true joint distribution of signals and parameters. It has as its θ^{th} row the unconditional probability that the signal is s and the parameter is θ. The marginal probability of the signals is given by the vector $P\mu$ and so the Bayesian update is

$$\mathcal{U}^B(\mu, P) := D_{(\mu)}P^T D^{-1}_{(P\mu)}. \tag{21}$$

Verifying that this satisfies (20) we can do two calculations. The first verifies that this rule generates probability vectors as columns: $\mathbb{1}^T\mathcal{U}^B(\mu, P) = \mu^T P^T D^{-1}_{(P\mu)} = \mathbb{1}^T$. The second verifies that the updating is unbiased

$$\mathcal{U}^B(\mu, P)P\mu = D_{(\mu)}P^T D^{-1}_{(P\mu)}P\mu = D_{(\mu)}P^T\mathbb{1} = D_{(\mu)}\mathbb{1} = \mu.$$

The next result is a characterisation of all the updating functions $\mathcal{U}(\mu, P)$ that satisfy the unbiased/martingale property. It shows that any such unbiased updating rule can be interpreted as a Bayesian update with a misspecified experiment. That is an updating rule is unbiased for the experiment $P = \mathcal{E}$ at μ if and only if it is a Bayesian update but for another experiment $\mathcal{E}' = Q$) such that $P\mu = Q\mu$.

Proposition 6. The updating function $\mathcal{U}(\mu, P)$ is unbiased if and only if $\mathcal{U}(\mu, P) = \mathcal{U}^B(\mu, Q)$ for some $Q \in \Delta(S)^{|\Theta|}$ satisfying $P\mu = Q\mu$.

Proof. Let us suppose that $\mathcal{U}(\mu, P) = \mathcal{U}^B(\mu, Q)$ for some $Q \in \Delta(S)^{|\Theta|}$ satisfying $P\mu = Q\mu$. To show that $\mathcal{U}$ is unbiased, a substitution from (21) gives gives

$$\mathcal{U}(\mu, P)P\mu = \mathcal{U}^B(\mu, Q)P\mu = D_{(\mu)}Q^T D^{-1}_{(Q\mu)}P\mu.$$

But as $P\mu = Q\mu$ this then can be written as

$$\mathcal{U}(\mu, P)P\mu = D_{(\mu)}Q^T D^{-1}_{(Q\mu)}Q\mu = D_{(\mu)}Q^T\mathbb{1} = D_{(\mu)}\mathbb{1} = \mu.$$

Hence we have that $\mathcal{U}(\mu, P)P\mu = \mu$ which is the unbiased property for the updating.

Now suppose that $\mathcal{U}(\mu, P)$, the updating function, is unbiased. This means it must satisfy the equation (20). If this is rearranged we get

$$\begin{aligned} \mu &= \mathcal{U}(\mu, P)P\mu \\ D^{-1}_{(\mu)}\mu &= D^{-1}_{(\mu)}\mathcal{U}(\mu, P)P\mu \end{aligned}$$

$$\mathbb{1} = D_{(\mu)}^{-1}\mathcal{U}(\mu,P)D_{(P\mu)}D_{(P\mu)}^{-1}P\mu$$
$$\mathbb{1} = \underbrace{D_{(\mu)}^{-1}\mathcal{U}(\mu,P)D_{(P\mu)}}_{Q^T}\mathbb{1}$$

Hence there is a $|\Theta|\times|S|$ matrix $Q^T := D_{(\mu)}^{-1}\mathcal{U}(\mu,P)D_{(P\mu)}$ (dependent on μ and P) satisfying $Q \in \Delta(S)^{|\Theta|}$. Transposing this gives $Q = D_{(P\mu)}\mathcal{U}(\mu,P)^T D_{(\mu)}^{-1}$ and so

$$Q\mu = D_{(P\mu)}\mathcal{U}(\mu,P)^T D_{(\mu)}^{-1}\mu = D_{(P\mu)}\mathcal{U}(\mu,P)^T\mathbb{1} = D_{(P\mu)}\mathbb{1} = P\mu.$$

The definition of $Q = D_{(P\mu)}\mathcal{U}(\mu,P)^T D_{(\mu)}^{-1}$, then can be inverted and transposed to write the updating function explicitly: $\mathcal{U}(\mu,P) = D_{(\mu)}Q^T D_{(P\mu)}^{-1}$. We have already shown that $Q\mu = P\mu$ and so $\mathcal{U}(\mu,P) = D_{(\mu)}Q^T D_{(Q\mu)}^{-1}$. A comparison with (21) then gives us that $\mathcal{U}(\mu,P) = \mathcal{U}^B(\mu,Q)$ and so the updating is a Bayesian update with the experiment $\mathcal{E} = Q$ which is what we wanted to show. □

8. Extensions and Applications

The assumption that the range of the updated beliefs is the whole probability simplex $\Delta(\Theta)$, made here, is inconsistent with several important models of non-Bayesian belief updating. For example, if the individual has limited memory or mental capacities, one might want to consider updating procedures that generate one of a finite set of probability measures. Thus the range of the updating function is finite. Models of learning where updated beliefs are discrete and finite, such as Hellman and Cover (1970), Dow (1991), and Wilson (2014), pose a considerable challenge for divisibility. It is not clear whether it is ever possible to satisfy divisibility of belief revision in such a setting.

There are other models of belief revision where the updates themselves are random. Thus the updating function maps an experiment and an initial belief to a profile of probability measures over updated beliefs. One example of this random updating is Rabin and Schrag (1999). It seems conceivable that a generalised notion of divisibility might apply in this setting.

References

Aczél, D., and J. Dhombres (1989): *Functional Equations in Several Variables*. Cambridge University Press, Cambridge, UK, second edn.

Aczél, D., and M. Hosszú (1956): "On Transformations with Several Parameters and Operations in Multidimensional Spaces," *Acta Math. Acad. Sci. Hungar.*, 6, 327–338.

Angrisani, M., A. Guarino, P. Jehiel, and T. Kitagawa (2017): "Information Redundancy Neglect Versus Overconfidence: A Social Learning Experiment," Cemmap working paper, UCL.

Berg, L. (1993): "The Local Structure of the Solutions of the Multidimensional Translation Equation," *Aequationes Mathematicae*, 46, 164–173.

Bohren, A., and D. Hauser (2017): "Bounded Rationality and Learning, A Framework and a Robustness Result," *under review*, pp. 349–374.

Brunnermeier, M. K. (2009): "Deciphering the Liquidity and Credit Crunch 2007–2008," *Journal of Economic Perspectives*, 23(1), 77–100.

Dawid, A. P. (1984): "Present position and potential developments: Some personal views: Statistical theory: The prequential approach," *Journal of the Royal Statistical Society. Series A (General)*, 147(2), 278–292.

DeGroot, M. H. (1970): *Optimal Statistical Decisions*. McGraw-Hill, New York.

Diaconis, P., and D. Freedman (1986): "On the Consistency of Bayes Estimates," *The Annals of Statistics*, 14(1), 1–26.

Dow, J. (1991): "Search decisions with limited memory," *The Review of Economic Studies*, 58(1), 1–14.

Epstein, L. G., J. Noor, and A. Sandroni (2010): "Non-Bayesian Learning," *The B.E. Journal of Theoretical Economics*, 10(1).

Epstein, L. G., and M. Schneider (2003): "Recursive Multiple-Priors," *Journal of Economic Theory*, 113(1), 1–31.

Epstein, L. G., and S. E. Zin (1989): "Substitution, Risk Aversion, and the Temporal Behavior of Consumption and Asset Returns: A Theoretical Framework," *Econometrica*, 57(4), 937–969.

Gilboa, I., and D. Schmeidler (1993): "Updating ambiguous beliefs," *Journal of economic theory*, 59(1), 33–49.

Hagmann, D., and G. Loewenstein (2017): "Persuasion with Motivated Beliefs," Carnegie Mellon University.

Hanany, E., and P. Klibanoff (2009): "Updating Ambiguity Averse Preferences," *The B.E. Journal of Theoretical Economics*, 9, 291–302.

Hellman, M. E., and T. M. Cover (1970): "Learning with finite memory," *The Annals of Mathematical Statistics*, pp. 765–782.

Horn, R. A., and C. R. Johnson (1985): *Matrix Analysis*. Cambridge University Press, Cambridge, UK.

Kallenberg, O. (2002): *Foundations of Modern Probability*. Springer Verlag, New York, second edn.

Lehrer, E., and R. Teper (2015): "Who is a Bayesian," Discussion paper, Working paper.

Levy, G., and R. Razin (2017): "Combining Forecasts: Why Decision Makers Neglect Correlation," Mimeo.

Moszner, Z. (1995): "General Theory of the Translation Equation," *Aequationes Mathematicae*, 50, 17–37.

Ortoleva, P. (2012): "Modeling the change of paradigm: Non-Bayesian reactions to unexpected news," *American Economic Review*, 102(6), 2410–36.

Rabin, M., and J. L. Schrag (1999): "First Impressions Matter: A Model of Confirmatory Bias," *The Quarterly Journal of Economics*, 114(1), 37–82.

Shannon, C. E., and W. Weaver (1949): *The Mathematical Theory of Communication*. University of Illinois Press, Urbana and Chicago.

Torgersen, E. (1991): *Comparison of Statistical Experiments*. Cambridge University Press, Cambridge, UK.

Tverberg, H. (1958): "A New Derivation of the Information Function," *Mathematica Scandinavica*, 6, 297–298.

Wilson, A. (2014): "Bounded memory and biases in information processing," *Econometrica*, 82(6), 2257–2294.

Zhao, C. (2016): "Pseudo-Bayesian Updating," mimeo.

Appendix

The Calculations for the Example in Section 1.1

Recall that $\mu = 1/2$ and:

$$\tilde{\mu}_2 = (1-\lambda)\tfrac{1}{2} + \lambda\frac{\alpha^2}{\beta^2+\alpha^2},$$

$$\mu_1 = (1-\lambda)\tfrac{1}{2} + \lambda\frac{\alpha}{\beta+\alpha}, \qquad \mu_2 = (1-\lambda)\mu_1 + \lambda\frac{\alpha\mu_1}{(1-\mu_1)\beta+\mu_1\alpha}.$$

We now take this expression for $\tilde{\mu}_2$ and make a sequence of substitutions. First for μ_1 and then for μ_2. A final re-arranging then gives the displayed equation.

$$\begin{aligned}
\tilde{\mu}_2 &= (1-\lambda)\tfrac{1}{2} + \lambda\frac{\alpha^2}{\beta^2+\alpha^2} \\
&= \mu_1 + \lambda\frac{\alpha^2}{\beta^2+\alpha^2} - \lambda\frac{\alpha}{\beta+\alpha} \\
&= \mu_2 + \lambda\frac{\alpha^2}{\beta^2+\alpha^2} - \lambda\frac{\alpha}{\beta+\alpha} + \lambda\mu_1 - \lambda\frac{\alpha\mu_1}{(1-\mu_1)\beta+\mu_1\alpha} \\
\frac{\tilde{\mu}_2-\mu_2}{\lambda} &= -\frac{\alpha\beta(\beta-\alpha)}{(\beta^2+\alpha^2)(\beta+\alpha)} + \frac{(\beta-\alpha)\mu_1(1-\mu_1)}{(1-\mu_1)\beta+\mu_1\alpha}
\end{aligned}$$

$$\frac{\tilde{\mu}_2 - \mu_2}{\lambda(\beta - \alpha)} = -\frac{\alpha\beta}{(\beta^2 + \alpha^2)(\beta + \alpha)} + \frac{1}{\frac{\beta}{\mu_1} + \frac{\alpha}{1-\mu_1}}$$

The final term above is an increasing function of μ_1 when $\mu_1 \in [0, 1/2]$. The RHS is positive when $\mu_1 = 1/2$ and negative when $\mu_1 = 0$. As λ increases from zero, so μ_1 decreases from $\frac{1}{2}$ to zero. Thus there exists a unique value $\bar{\lambda}$ such that $\tilde{\mu}_2 > \mu_2$ if an only if $\lambda < \bar{\lambda}$. But $\tilde{\mu}_2 = \mu_2$ when $\lambda = 1$ so we know $\tilde{\mu}_2 > \mu_2$ if an only if $\lambda < 1$.

Axiom on Order Reversal and its Implications

To describe this formally we consider a subset of signals: $A \subset S$. Then, we posit a pair of two-step procedures for learning the signal s: The first procedure begins with an experiment where the agent is told the signal s only if the realised s is in the set A. But if the realised signal s is not in the set A, then the agent is just told $s \notin A$. If they were told $s \notin A$ in the first experiment, then a second experiment is run where the agent learns the value of $s \in S \setminus A$.

The second procedure reverses this order. First the agent runs an experiment where they learn the value of s only if $s \in S \setminus A$. If the signal is not in this set then they are told just $s \in A$. If they were told $s \in A$ at the first stage, then a second experiment is run where they learn the value of $s \in A$. In Axiom 6 below we requires that both of these processes for learning the signal s result in the same terminal belief profile. We do not, however, require that either of these belief revision procedures has a terminal belief profile that is the same as the one-step process above.

Some additional notation is necessary for the formal statement of this Axiom. The set $A := \{1, 2, \ldots, m\}$ and $S \setminus A := \{m+1, \ldots, n\}$. The fact that we restrict A to only consist of the first m signals is unimportant, because of the symmetry axiom this will apply to all non-empty subsets of signals A. As A has m elements, the update after the first experiment of the first procedure is described by the function U_{m+1}. Whereas the update after the first experiment of the second procedure is described by the function U_{n-m+1}, because $S \setminus A$ has $n - m$ elements. In the first experiment of the first procedure the signals have the probabilities $p_A^\theta := (p_1^\theta, \ldots, p_m^\theta, q_A^\theta) \in \Delta^o(m+1)$, where $q_A^\theta = 1 - \sum_{k=1}^m p_k^\theta$. Similarly in the first experiment in the second procedure the signals have the probabilities $p_{S\setminus A}^\theta = (q_{S\setminus A}^\theta, p_{m+1}^\theta, \ldots, p_n^\theta) \in \Delta(n-m+1)$, where $q_{S\setminus A}^\theta = 1 - \sum_{k=m+1}^n p_k^\theta$. Thus the first experiment in the first procedure results in the profile of updated beliefs $U_{m+1}(\mu, (p_A^\theta)_{\theta\in\Theta})$ and the first experiment in the second procedure has the updated beliefs $U_{n-m+1}(\mu, (p_{S\setminus A}^\theta)_{\theta\in\Theta})$.

Axiom 6 (Order Independence)**.** For all $\mathcal{E} = (S, (p^\theta)_{\theta\in\Theta})$, $\mu \in \Delta(\Theta)$, any $m \leq n$, and $n \geq 3$.

$$\left[U_{m+1}^1\Big(\mu, (p_A^\theta)_{\theta\in\Theta}\Big), \ldots, U_{m+1}^m\Big(\mu, (p_A^\theta)_{\theta\in\Theta}\Big), U_{n-m}\left(U_{m+1}^{m+1}\Big(\mu, (p_A^\theta)_{\theta\in\Theta}\Big), \left(\frac{\tilde{p}_{S\setminus A}^\theta}{q_A^\theta}\right)_{\theta\in\Theta}\right)\right]$$

$$\equiv$$

$$\left[U_m\left(U^1_{n-m+1}\left(\mu,(p^\theta_{S\setminus A})_{\theta\in\Theta}\right),\left(\frac{\tilde p^\theta_A}{q^\theta_{S\setminus A}}\right)_{\theta\in\Theta}\right),U^2_{n-m+1}\left(\mu,(p^\theta_{S\setminus A})_{\theta\in\Theta}\right),\dots,U^{n-m+1}_{n-m+1}\left(\mu,(p^\theta_{S\setminus A})_{\theta\in\Theta}\right)\right]$$

Where: $\tilde p^\theta_A := (p^\theta_1,\dots,p^\theta_m)$, $\tilde p^\theta_{S\setminus A} := (p^\theta_{m+1},\dots,p^\theta_n)$.

Axiom 6 does not impose the condition that either of the terms in this equality is equal to the profile of beliefs in the one-off belief revision process: $U_n(\mu,(p^\theta)_{\theta\in\Theta})$. Axiom 4 says that when $m=1$ both of these terms are equal to $U_n(\mu,(p^\theta)_{\theta\in\Theta})$. Then, an iterated application of Axiom 4 will imply they are equal for all m. Thus Axiom 6 is implied by Axiom 4 and symmetry, but potentially this axiom is weaker than divisibility.

To see that this is not actually the case, and Axiom 6 implies Axiom 4, consider this condition when $m=1$. In the first procedure, where $s=1$ or $s\neq 1$ is learned first the terminal belief profile is:

$$\left[U^1_2\left(\mu,(p^\theta_1,1-p^\theta_1)_{\theta\in\Theta}\right),U_{n-1}\left(U^2_2\left(\mu,(p^\theta_1,1-p^\theta_1)_{\theta\in\Theta}\right),\left(\frac{p^\theta_{-1}}{1-p^\theta_1}\right)_{\theta\in\Theta}\right)\right]$$

Or, recalling the definition (2),

$$\left[u\left(\mu,(p^\theta_1)_{\theta\in\Theta}\right),U_{n-1}\left(u\left(\mu,(1-p^\theta_1)_{\theta\in\Theta}\right),\left(\frac{p^\theta_{-1}}{1-p^\theta_1}\right)_{\theta\in\Theta}\right)\right].$$

However, (when $m=1$) the terminal belief profile after the second procedure for learning the signal is $U_n(\mu,(p^\theta)_{\theta\in\Theta})$. This is because under this procedure at the first stage the agent learns s if $s\geq 2$ and otherwise learns that $s<2$. But this implies that the first stage of the second procedure completely reveals s and there is no additional learning at the second stage. Thus when $m=1$ Axiom 6 implies

$$U_n(\mu,(p^\theta)_{\theta\in\Theta})=\left[u\left(\mu,(p^\theta_1)_{\theta\in\Theta}\right),U_{n-1}\left(u\left(\mu,(1-p^\theta_1)_{\theta\in\Theta}\right),\left(\frac{p^\theta_{-1}}{1-p^\theta_1}\right)_{\theta\in\Theta}\right)\right],$$

which is precisely the condition in Axiom 4.

Proof of Proposition 2

Proof. The updating satisfies the axioms required for Proposition 1 to hold, hence it is characterised by a homeomorphism $F:\Delta(\Theta)\to\Delta(\Theta)$. We can, therefore, define the stochastic process followed by the shadow beliefs, $\{\tilde\mu^t\}_{t=0}^\infty$, where $\tilde\mu^t := F(\mu^t)$. Applying the updating rule (8) in period t with the signal s^t to the initial beliefs μ^t, we have that the updated beliefs satisfy

$$F(\mu^{t+1})=F(u(\mu^t,(p^\theta_{s^t})_{\theta\in\Theta}))=\left(\frac{F_1(\mu^t)p^1_{s^t}}{\sum_{\theta\in\Theta}F_\theta(\mu^t)p^\theta_{s^t}},\dots,\frac{F_{|\Theta|}(\mu^t)p^{|\Theta|}_{s^t}}{\sum_{\theta\in\Theta}F_\theta(\mu^t)p^\theta_{s^t}}\right).$$

Substitution for $\tilde{\mu}^t$ then gives

$$\tilde{\mu}^{t+1} = \left(\frac{\tilde{\mu}^t_1 p^1_{s^t}}{\sum_{\theta\in\Theta} \tilde{\mu}^t_\theta p^\theta_{s^t}}, \ldots, \frac{\tilde{\mu}^t_{|\Theta|} p^{|\Theta|}_{s^t}}{\sum_{\theta\in\Theta} \tilde{\mu}^t_\theta p^\theta_{s^t}} \right).$$

Applying the above we can, then, write

$$\frac{\tilde{\mu}^t_\theta}{\tilde{\mu}^t_{\theta'}} = \frac{\tilde{\mu}^{t-1}_\theta}{\tilde{\mu}^{t-1}_{\theta'}} \frac{p^\theta_{s^{t-1}}}{p^{\theta'}_{s^{t-1}}}.$$

This is the expression for the Bayesian updating of the shadow beliefs. Iterating this relation and taking logarithms we get

$$\frac{1}{t} \ln \frac{\tilde{\mu}^t_\theta}{\tilde{\mu}^t_{\theta'}} = \frac{1}{t} \ln \frac{\tilde{\mu}^0_\theta}{\tilde{\mu}^0_{\theta'}} + \frac{1}{t} \sum_{\tau=0}^{t-1} \ln \frac{p^\theta_{s^\tau}}{p^{\theta'}_{s^\tau}}, \qquad\qquad \tilde{\mu}^0_{\theta'} > 0.$$

When the parameter is θ, the terms in the above summation are independently and identically distributed with the expectation $H(p^\theta \| p^{\theta'}) = \sum_s p^\theta_s \ln \frac{p^\theta_s}{p^{\theta'}_s} \geq 0$. This is the relative entropy of the measures p^θ and $p^{\theta'}$. By the Strong Law of Large Numbers (Kallenberg (2002) p.73)

$$\lim_{t\to\infty} \frac{1}{t} \ln \frac{\tilde{\mu}^t_\theta}{\tilde{\mu}^t_{\theta'}} \to H(p^\theta \| p^{\theta'}), \qquad \forall \theta' \neq \theta;$$

$\mathbb{P}^\theta$ almost surely. For θ' with $H(p^\theta \| p^{\theta'}) > 0$ this implies $\tilde{\mu}^t_{\theta'} \to 0$ $\mathbb{P}^\theta$ almost surely. And $\frac{\tilde{\mu}^t_\theta}{\tilde{\mu}^t_{\theta'}} = \frac{\tilde{\mu}^0_\theta}{\tilde{\mu}^0_{\theta'}}$ for θ' with $H(p^\theta \| p^{\theta'}) = 0$. Thus $\tilde{\mu}^t$ converges almost surely to $\tilde{\mu}^\infty \in \Delta(\Theta)$. As F is a homeomorphism this implies $\mu^t = F^{-1}(\tilde{\mu}^t)$ converges almost surely to $\mu^\infty = F^{-1}(\tilde{\mu}^\infty)$.

When the assumption $p^\theta \neq p^{\theta'}$ for all θ' holds then $H(p^\theta \| p^{\theta'}) > 0$ for all θ' and $\tilde{\mu}^\infty = e_\theta$. By (12) this implies that $\mu^\infty = F^{-1}(\tilde{\mu}^\infty) = e_\theta$. □

Proof of Proposition 3

Proof. We begin by establishing part (i) of the result. Re-arranging (14) we get a relation for $u_\theta(\mu, p^s)$, the updated belief in the parameter θ conditional on the signal s,

$$f \circ u_\theta(\mu, p^s) = \frac{f(\mu_\theta) p^\theta_s}{f(\mu_\theta) p^\theta_s + (1 - f(\mu_\theta)) p^{\theta'}_s}.$$

This is the shadow posterior on θ. Now we calculate the shadow posterior odds ratio. Then take its expectation conditional on the parameter θ, that is

$$E^\theta \left(\frac{1 - f \circ u_\theta(\mu, p^s)}{f \circ u_\theta(\mu, p^s)} \right) = \sum_s p^\theta_s \frac{(1 - f(\mu_\theta)) p^{\theta'}_s}{f(\mu_\theta) p^\theta_s} = \frac{1 - f(\mu_\theta)}{f(\mu_\theta)}.$$

Adding unity to the extremes of this equality gives

$$E^\theta \left(\frac{1}{f \circ u_\theta(\mu, p^s)} \right) = \frac{1}{f(\mu_\theta)}.$$

The function $\frac{1}{f(.)}$ is assumed to be convex on an interval of values containing the points $\{u_\theta(\mu, p^s) : s \in S\}$. Therefore, by Jensen's inequality

$$\frac{1}{f \circ E^\theta(u_\theta(\mu, p^s))} \leq E^\theta \left(\frac{1}{f \circ u_\theta(\mu, p^s)} \right) = \frac{1}{f(\mu_\theta)}.$$

When $f(.)$ is increasing the extremes of this inequality imply that $\mu_\theta \leq E^\theta(u_\theta(\mu, p^s))$. This establishes the first part of the result. Part (ii) is established by observing that the final inequality is reversed when concavity replaces convexity.

Finally, we must show that $\frac{1}{f(.)}$ cannot be concave on the open interval $(0, x)$, for any $x > 0$. Suppose it were concave on such an interval for some $x \in (0, 1)$. Then for $\varepsilon < x$ and any $\lambda \in [0, 1]$

$$\frac{1}{f(\lambda\varepsilon + (1-\lambda)x)} \geq \lambda \frac{1}{f(\varepsilon)} + (1-\lambda)\frac{1}{f(x)},$$

But as $\varepsilon \to 0$ the RHS of the first of these inequalities converges to infinity (as $f(\varepsilon) \to 0$). Thus $f((1-\lambda)x) = 0$ for all $\lambda \in (0, 1)$ which contradicts the fact that $f(.)$ is strictly increasing. □

Proof of Proposition 4

Proof. A differentiable, strictly increasing map $f : [0, 1] \to [0, 1]$ determines the divisible updating $u_\theta(\mu, p^s)$ as in (14). We define the (differentiable and strictly increasing) function $\psi : \mathbb{R} \to \mathbb{R}$ as

$$\psi(\lambda) = \ln f\left(\frac{e^\lambda}{1+e^\lambda}\right) - \ln\left[1 - f\left(\frac{e^\lambda}{1+e^\lambda}\right)\right], \qquad \lambda \in \mathbb{R}.$$

Also, note for later that

$$\psi'(\lambda) = f'(x)\frac{x}{f(x)}\frac{1-x}{1-f(x)}, \qquad \text{where} \qquad x := \frac{e^\lambda}{1+e^\lambda}. \tag{22}$$

If we define $\lambda'_s := \ln \frac{u_\theta(\mu, p_s)}{1-u_\theta(\mu, p_s)}$ and $\lambda := \ln \frac{\mu_\theta}{1-\mu_\theta}$, then (14) implies that

$$\psi(\lambda'_s) = \psi(\lambda) + \ln \frac{p_s^\theta}{p_s^{\theta'}}.$$

The function ψ is invertible, so this allows us to write

$$\lambda'_s = \ln \frac{u_\theta(\mu, p_s)}{1-u_\theta(\mu, p_s)} = \psi^{-1}\left(\psi(\lambda) + \ln \frac{p_s^\theta}{p_s^{\theta'}}\right). \tag{23}$$

Now we calculate the variance of the updated beliefs log likelihood ratio. This is

$$\mathrm{Var}\left[\ln \frac{\mathrm{u}_\theta(\mu, \mathrm{p_s})}{1-\mathrm{u}_\theta(\mu, \mathrm{p_s})}\right] = \frac{1}{2}\sum_{\mathrm{s,s'\in S}} \pi_\mathrm{s}\pi_{\mathrm{s'}}\left(\ln \frac{\mathrm{u}_\theta(\mu, \mathrm{p_s})}{1-\mathrm{u}_\theta(\mu, \mathrm{p_s})} - \ln \frac{\mathrm{u}_\theta(\mu, \mathrm{p_{s'}})}{1-\mathrm{u}_\theta(\mu, \mathrm{p_{s'}})}\right)^2.$$

Here $\pi_s = \sum_\theta \mu^\theta p_s^\theta$ is defined to be the unconditional probability of the signal s in the experiment $\mathcal{E}$. A substitution from (23) then gives

$$\mathrm{Var}\left[\ln \frac{\mathrm{u}_\theta(\mu, \mathrm{p_s})}{1-\mathrm{u}_\theta(\mu, \mathrm{p_s})}\right] = \frac{1}{2}\sum_{s,s' \in S} \pi_s \pi_{s'} \left(\psi^{-1}\left(\psi(\lambda) + \ln \frac{p_s^\theta}{p_s^{\theta'}}\right) - \psi^{-1}\left(\psi(\lambda) + \ln \frac{p_{s'}^\theta}{p_{s'}^{\theta'}}\right)\right)^2.$$

As we have assumed the function f is continuously differentiable we can apply the intermediate value theorem to the function ψ^{-1}. Hence,

$$\psi^{-1}\left(\psi(\lambda) + \ln \frac{p_s^\theta}{p_s^{\theta'}}\right) - \psi^{-1}\left(\psi(\lambda) + \ln \frac{p_{s'}^\theta}{p_{s'}^{\theta'}}\right) = \frac{d\psi^{-1}(\tilde{\lambda})}{d\lambda}\left(\ln \tfrac{p_s^\theta}{p_s^{\theta'}} - \ln \tfrac{p_{s'}^\theta}{p_{s'}^{\theta'}}\right)$$

for some $\tilde{\lambda}$ satisfying $\min_s \ln \frac{p_s^\theta}{p_s^{\theta'}} \le \tilde{\lambda} - \psi(\lambda) \le \max_s \ln \frac{p_s^\theta}{p_s^{\theta'}}$. Let B denote this interval of potential values of $\tilde{\lambda}$, then if this calculation is substituted into the expression for the variance we can then get a lower bound on the variance

$$\begin{aligned}\mathrm{Var}\left[\ln \frac{\mathrm{u}_\theta(\mu, \mathrm{p_s})}{1-\mathrm{u}_\theta(\mu, \mathrm{p_s})}\right] &\ge \min_{\tilde{\lambda}\in B}\left[\frac{d\psi^{-1}(\tilde{\lambda})}{d\lambda}\right]^2 \frac{1}{2}\sum_{s,s' \in S} \pi_s \pi_{s'} \left(\ln \frac{p_s^\theta}{p_s^{\theta'}} - \ln \frac{p_{s'}^\theta}{p_{s'}^{\theta'}}\right)^2 \\ &= \min_{\tilde{\lambda}\in B}\left[\frac{d\psi^{-1}(\tilde{\lambda})}{d\lambda}\right]^2 \mathrm{Var}\left[\ln \frac{\mathrm{p}^\theta}{\mathrm{p}^{\theta'}}\right].\end{aligned}$$

An upper bound can be obtained in a similar way

$$\mathrm{Var}\left[\ln \frac{\mathrm{u}_\theta(\mu, \mathrm{p_s})}{1-\mathrm{u}_\theta(\mu, \mathrm{p_s})}\right] \le \max_{\tilde{\lambda}\in B}\left[\frac{d\psi^{-1}(\tilde{\lambda})}{d\lambda}\right]^2 \mathrm{Var}\left[\ln \frac{\mathrm{p}^\theta}{\mathrm{p}^{\theta'}}\right].$$

These inequalities imply that bounding the derivatives of ψ^{-1} will generate over and under reaction. As ψ and its inverse are strictly increasing functions with positive derivatives. The above inequalities imply that a sufficient condition for the updating to satisfy the condition for overreaction is $d\psi^{-1}(\tilde{\lambda})/d\lambda > 1$ for all $\tilde{\lambda}$ and a sufficient condition for under-reaction is $d\psi^{-1}(\tilde{\lambda})/d\lambda < 1$. The calculation of the derivative $d\psi/d\lambda$ in (22) then implies the sufficient conditions given in the Proposition. □