Divisible Updating

Martin Cripps

UCL

2018

Model and Notation

I study a model of updating of beliefs:

- Unknown parameter $\theta \in \{1, 2, \ldots, |\Theta|\} := \Theta$
- Initial Beliefs $\mu = (\mu^1, \ldots, \mu^{|\Theta|}) \in \Delta(\Theta)$
- Signals $s \in \{1, 2, \ldots, n\} = S$
- Statistical experiment $\mathcal{E} := ((p^\theta)_{\theta \in \Theta}) \in \Delta^o(S)^K$.
- $p^\theta = (p_1^\theta, \ldots, p_n^\theta) > 0$.

Model and Notation

I study a model of updating of beliefs:

- Unknown parameter $\theta \in \{1, 2, \dots, |\Theta|\} := \Theta$
- Initial Beliefs $\mu = (\mu^1, \dots, \mu^{|\Theta|}) \in \Delta(\Theta)$
- Signals $s \in \{1, 2, \dots, n\} = S$
- Statistical experiment $\mathcal{E} := ((p^\theta)_{\theta \in \Theta}) \in \Delta^o(S)^K$.
- $p^\theta = (p^\theta_1, \dots, p^\theta_n) > 0$.

Model and Notation

I study a model of updating of beliefs:

- Unknown parameter $\theta \in \{1, 2, \ldots, |\Theta|\} := \Theta$
- Initial Beliefs $\mu = (\mu^1, \ldots, \mu^{|\Theta|}) \in \Delta(\Theta)$
- Signals $s \in \{1, 2, \ldots, n\} = S$
- Statistical experiment $\mathcal{E} := ((p^\theta)_{\theta \in \Theta}) \in \Delta^o(S)^K$.
- $p^\theta = (p_1^\theta, \ldots, p_n^\theta) > 0$.

Model and Notation

I study a model of updating of beliefs:

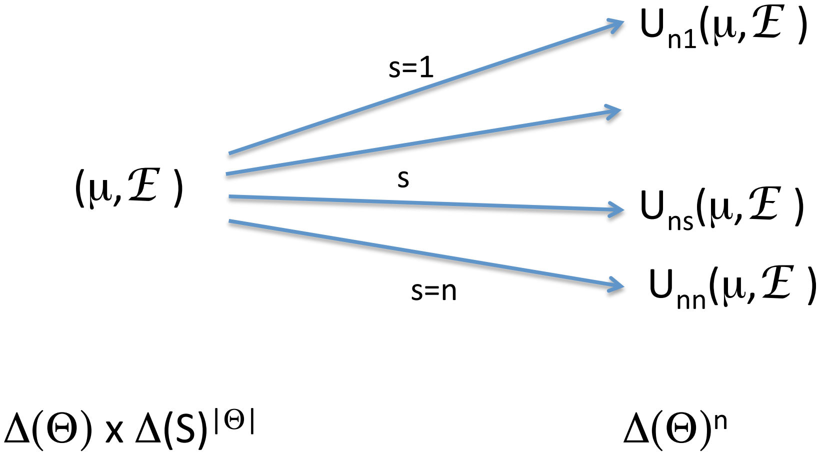
- Unknown parameter $\theta \in \{1, 2, \ldots, |\Theta|\} := \Theta$
- Initial Beliefs $\mu = (\mu^1, \ldots, \mu^{|\Theta|}) \in \Delta(\Theta)$
- Signals $s \in \{1, 2, \ldots, n\} = S$
- Statistical experiment $\mathcal{E} := ((p^\theta)_{\theta \in \Theta}) \in \Delta^o(S)^K$.
- $p^\theta = (p_1^\theta, \ldots, p_n^\theta) > 0$.

Model and Notation

I study a model of updating of beliefs:

- Unknown parameter $\theta \in \{1, 2, \ldots, |\Theta|\} := \Theta$
- Initial Beliefs $\mu = (\mu^1, \ldots, \mu^{|\Theta|}) \in \Delta(\Theta)$
- Signals $s \in \{1, 2, \ldots, n\} = S$
- Statistical experiment $\mathcal{E} := ((p^\theta)_{\theta \in \Theta}) \in \Delta^o(S)^K$.
- $p^\theta = (p_1^\theta, \ldots, p_n^\theta) > 0$.

The Updating Function



Updating Rule U_n

- U_n is a map from the beliefs and the experiment to a profile of updated beliefs:$U_n(\mu, p^1, \ldots, p^{|\Theta|}) = (U_{n1}, \ldots, U_{nn})$

$$U_n : \Delta(\Theta) \times \Delta^o(S)^K \to \Delta(\Theta)^n, \qquad n = 2, 3, \ldots .$$

- We will impose some conditions on the function U_n and see what updating rules are consistent with these.

Some Properties we might want U_n to have

1. No update if signals uninformative: $U_n(\mu, p, \ldots, p) = (\mu, \ldots, \mu)$, for all $p \in \Delta^o(S)$, $\mu \in \Delta(\Theta)$ and n.
2. The names of the signals do not matter—reorder the signals but don't change their probabilities and you just get a re-ordering of U_n.Symmetry
3. Divisibility — see later.
4. If there are only two signals, you can find an experiment that generates any updated belief you want for any one signal *and* updating is one to one. Non-Dogmatic

Some Properties we might want U_n to have

1. No update if signals uninformative: $U_n(\mu, p, \ldots, p) = (\mu, \ldots, \mu)$, for all $p \in \Delta^o(S)$, $\mu \in \Delta(\Theta)$ and n.
2. The names of the signals do not matter—reorder the signals but don't change their probabilities and you just get a re-ordering of U_n.Symmetry
3. Divisibility — see later.
4. If there are only two signals, you can find an experiment that generates any updated belief you want for any one signal *and* updating is one to one. Non-Dogmatic

Some Properties we might want U_n to have

1. No update if signals uninformative: $U_n(\mu, p, \ldots, p) = (\mu, \ldots, \mu)$, for all $p \in \Delta^o(S)$, $\mu \in \Delta(\Theta)$ and n.
2. The names of the signals do not matter—reorder the signals but don't change their probabilities and you just get a re-ordering of U_n.Symmetry
3. Divisibility — see later.
4. If there are only two signals, you can find an experiment that generates any updated belief you want for any one signal *and* updating is one to one. Non-Dogmatic

Some Properties we might want U_n to have

1. No update if signals uninformative: $U_n(\mu, p, \ldots, p) = (\mu, \ldots, \mu)$, for all $p \in \Delta^o(S)$, $\mu \in \Delta(\Theta)$ and n.
2. The names of the signals do not matter—reorder the signals but don't change their probabilities and you just get a re-ordering of U_n.Symmetry
3. Divisibility — see later.
4. If there are only two signals, you can find an experiment that generates any updated belief you want for any one signal *and* updating is one to one. Non-Dogmatic

Divisibility

1. Typically information/signals comes in bundles: the birthday present is small but it has expensive gift wrapping.
2. We can process this information in several ways all at once —by treating the bundle as a signal from a joint distribution.
3. Or we can process this information in stages —That is, to update beliefs once using the first piece of information and its distribution. And then to update these intermediate beliefs a second time using the second piece of information and its conditional distribution given the first piece of information.
4. Divisibility says that both of these processes generate the same profile of beliefs

Divisibility: Why?

1. If updating is not divisible — one updating rule does not specify an individual's beliefs. We need to know when the updating rule is being applied.
2. Is a property that is easy to explain to subjects—most would agree that it is normatively reasonable.
3. Ensures a dynamic consistency of beliefs.
4. In a dynamic setting is that it allows one summary statistic — current beliefs. If beliefs are not divisible then in a dynamic setting may need to keep track of more things.
5. It to allows one studied departure from Bayes: Angrisani, Guarino, Jehiel, and Kitagawa (2017).

Divisibility: Why?

1. If updating is not divisible — one updating rule does not specify an individual's beliefs. We need to know when the updating rule is being applied.
2. Is a property that is easy to explain to subjects—most would agree that it is normatively reasonable.
3. Ensures a dynamic consistency of beliefs.
4. In a dynamic setting is that it allows one summary statistic — current beliefs. If beliefs are not divisible then in a dynamic setting may need to keep track of more things.
5. It to allows one studied departure from Bayes: Angrisani, Guarino, Jehiel, and Kitagawa (2017).

Divisibility: Why?

1. If updating is not divisible — one updating rule does not specify an individual's beliefs. We need to know when the updating rule is being applied.
2. Is a property that is easy to explain to subjects—most would agree that it is normatively reasonable.
3. Ensures a dynamic consistency of beliefs.
4. In a dynamic setting is that it allows one summary statistic — current beliefs. If beliefs are not divisible then in a dynamic setting may need to keep track of more things.
5. It to allows one studied departure from Bayes: Angrisani, Guarino, Jehiel, and Kitagawa (2017).

Divisibility: Why?

1. If updating is not divisible — one updating rule does not specify an individual's beliefs. We need to know when the updating rule is being applied.
2. Is a property that is easy to explain to subjects—most would agree that it is normatively reasonable.
3. Ensures a dynamic consistency of beliefs.
4. In a dynamic setting is that it allows one summary statistic — current beliefs. If beliefs are not divisible then in a dynamic setting may need to keep track of more things.
5. It to allows one studied departure from Bayes: Angrisani, Guarino, Jehiel, and Kitagawa (2017).

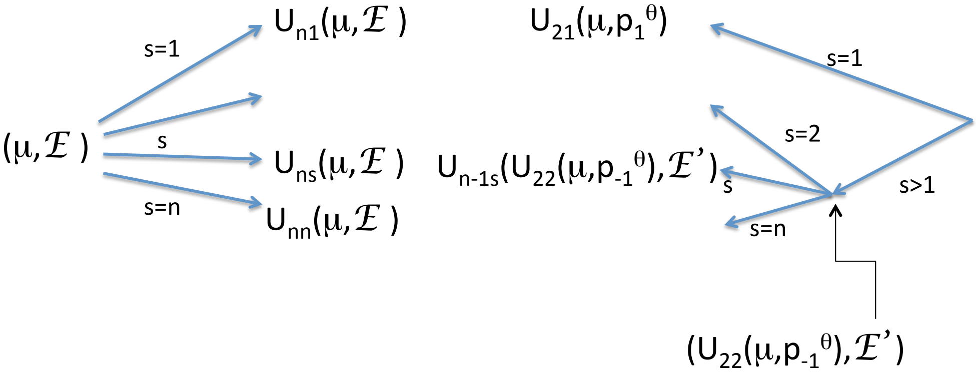
Divisibility: Why?

1. If updating is not divisible — one updating rule does not specify an individual's beliefs. We need to know when the updating rule is being applied.
2. Is a property that is easy to explain to subjects—most would agree that it is normatively reasonable.
3. Ensures a dynamic consistency of beliefs.
4. In a dynamic setting is that it allows one summary statistic — current beliefs. If beliefs are not divisible then in a dynamic setting may need to keep track of more things.
5. It to allows one studied departure from Bayes: Angrisani, Guarino, Jehiel, and Kitagawa (2017).

Some of the Literature

- Alternatives/Improvements on Bayesian updating that generate interesting properties (overconfidence, biases, correlation neglect, interesting biases): Rabin and Schrag (1999), Ortoleva (2012), Angrisani, Guarino, Jehiel, and Kitagawa (2017), Levy and Razin (2017), Brunnermeier (2009), Bohren and Hauser (2017), Epstein, Noor, and Sandroni (2010)
- Dynamically consistent preferences, exchangability of actions: Epstein and Zin (1989), Epstein and Schneider (2003), Ahn, Echenique, and Saito (2018) .
- Divisibility: Gilboa and Schmeidler (1993) called "commutativity".
- Hanany and Klibanoff (2009), show that a "reweighted Bayesian update" satisfies divisibility.
- Zhao (2016) — order independence property.
- Statistics Dawid (1984),

Divisibility



Divisibility: Formally

$$U_n(\mu, \mathcal{E}) \equiv \left[U_{21}(\mu, p_1), U_{n-1}\big(U_{22}(\mu, \mathbf{1} - p_1), \mathcal{E}'\big)\right].$$

$p_1 := (p_1^\theta : \theta \in \Theta)$. Here $\mathcal{E}'$ is the conditional experiment with signals $s = 2, 3, \ldots, n$.

$$\mathcal{E}' := \left(\frac{p_{-1}^\theta}{1 - p_1^\theta}\right)_{\theta \in \Theta}$$

An Example of Non-Divisible Updating

1. Arrival process: Good state a bus will arrive in period $t \geq 0$ with probability $(1-\alpha)\alpha^t$; Bad state $(1-\beta)\beta^t$ $(\alpha < \beta)$.
2. $\mu = 1/2$ that the state is good.
3. If no bus arrives in period $t = 0$, then Bayesian revision gives $\mu' = \frac{\alpha}{\beta+\alpha}$.
4. Epstein, Noor, and Sandroni (2010), Hagmann and Loewenstein (2017)

$$\mu_1 = (1-\lambda)\tfrac{1}{2} + \lambda \frac{\alpha}{\beta+\alpha}, \qquad \lambda \geq 0. \tag{1}$$

5. In $t = 2$ revised beliefs would be

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

6. If arrived in $t = 2$ and just did one big update

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

An Example of Non-Divisible Updating

1. Arrival process: Good state a bus will arrive in period $t \geq 0$ with probability $(1-\alpha)\alpha^t$; Bad state $(1-\beta)\beta^t$ $(\alpha < \beta)$.
2. $\mu = 1/2$ that the state is good.
3. If no bus arrives in period $t = 0$, then Bayesian revision gives $\mu' = \frac{\alpha}{\beta+\alpha}$.
4. Epstein, Noor, and Sandroni (2010), Hagmann and Loewenstein (2017)

$$\mu_1 = (1-\lambda)\tfrac{1}{2} + \lambda\frac{\alpha}{\beta+\alpha}, \qquad \lambda \geq 0. \tag{1}$$

5. In $t = 2$ revised beliefs would be

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

6. If arrived in $t = 2$ and just did one big update

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

An Example of Non-Divisible Updating

1. Arrival process: Good state a bus will arrive in period $t \geq 0$ with probability $(1-\alpha)\alpha^t$; Bad state $(1-\beta)\beta^t$ $(\alpha < \beta)$.
2. $\mu = 1/2$ that the state is good.
3. If no bus arrives in period $t = 0$, then Bayesian revision gives $\mu' = \frac{\alpha}{\beta+\alpha}$.
4. Epstein, Noor, and Sandroni (2010), Hagmann and Loewenstein (2017)

$$\mu_1 = (1-\lambda)\tfrac{1}{2} + \lambda\frac{\alpha}{\beta+\alpha}, \qquad \lambda \geq 0. \tag{1}$$

5. In $t = 2$ revised beliefs would be

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

6. If arrived in $t = 2$ and just did one big update

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

An Example of Non-Divisible Updating

1. Arrival process: Good state a bus will arrive in period $t \geq 0$ with probability $(1-\alpha)\alpha^t$; Bad state $(1-\beta)\beta^t$ $(\alpha < \beta)$.
2. $\mu = 1/2$ that the state is good.
3. If no bus arrives in period $t = 0$, then Bayesian revision gives $\mu' = \frac{\alpha}{\beta+\alpha}$.
4. Epstein, Noor, and Sandroni (2010), Hagmann and Loewenstein (2017)

$$\mu_1 = (1-\lambda)\tfrac{1}{2} + \lambda\frac{\alpha}{\beta+\alpha}, \qquad \lambda \geq 0. \tag{1}$$

5. In $t = 2$ revised beliefs would be

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

6. If arrived in $t = 2$ and just did one big update

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

An Example of Non-Divisible Updating

1. Arrival process: Good state a bus will arrive in period $t \geq 0$ with probability $(1-\alpha)\alpha^t$; Bad state $(1-\beta)\beta^t$ ($\alpha < \beta$).
2. $\mu = 1/2$ that the state is good.
3. If no bus arrives in period $t = 0$, then Bayesian revision gives $\mu' = \frac{\alpha}{\beta+\alpha}$.
4. Epstein, Noor, and Sandroni (2010), Hagmann and Loewenstein (2017)

$$\mu_1 = (1-\lambda)\tfrac{1}{2} + \lambda\frac{\alpha}{\beta+\alpha}, \qquad \lambda \geq 0. \tag{1}$$

5. In $t = 2$ revised beliefs would be

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

6. If arrived in $t = 2$ and just did one big update

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

Examples of Divisible Updating

1. Weighted Bayes $\mu_1 = \frac{\alpha^x \mu}{\mu\alpha^x + (1-\mu)\beta^x}$

$$\frac{\mu_2}{1-\mu_2} = \frac{\alpha^x}{\beta^x}\frac{\mu_1}{1-\mu_1} = \frac{(\alpha^2)^x}{(\beta^2)^x}\frac{\mu_0}{1-\mu_0}$$

2. Trigonometric $\tan\frac{\pi}{2}\mu_1 = \sqrt{\frac{\alpha}{\beta}}\tan\frac{\pi}{2}\mu$

$$\tan\frac{\pi}{2}\mu_2 = \sqrt{\frac{\alpha}{\beta}}\tan\frac{\pi}{2}\mu_1 = \sqrt{\frac{\alpha^2}{\beta^2}}\tan\frac{\pi}{2}\mu_0$$

Examples of Divisible Updating

1. Weighted Bayes $\mu_1 = \frac{\alpha^x \mu}{\mu\alpha^x + (1-\mu)\beta^x}$

$$\frac{\mu_2}{1-\mu_2} = \frac{\alpha^x}{\beta^x}\frac{\mu_1}{1-\mu_1} = \frac{(\alpha^2)^x}{(\beta^2)^x}\frac{\mu_0}{1-\mu_0}$$

2. Trigonometric $\tan \frac{\pi}{2}\mu_1 = \sqrt{\frac{\alpha}{\beta}} \tan \frac{\pi}{2}\mu$

$$\tan \frac{\pi}{2}\mu_2 = \sqrt{\frac{\alpha}{\beta}} \tan \frac{\pi}{2}\mu_1 = \sqrt{\frac{\alpha^2}{\beta^2}} \tan \frac{\pi}{2}\mu_0$$

The Characterisation Result

The updating satisfies (uninformativeness, symmetry, non-dogmatism, divisibility), iff there exists a homeomorphism $F : \Delta(\Theta) \to \Delta(\Theta)$ such that the updating satisfies

$$\begin{array}{c} \text{Beliefs} \\ \big\downarrow F \\ \text{Shadow Prior} \end{array}$$

The Characterisation Result

The updating satisfies (uninformativeness, symmetry, non-dogmatism, divisibility), iff there exists a homeomorphism $F : \Delta(\Theta) \to \Delta(\Theta)$ such that the updating satisfies

$$\begin{array}{ccc} \text{Beliefs} & & \\ \downarrow F & & \\ \text{Shadow Prior} & \xrightarrow{\text{Bayes Updating}} & \text{Shadow Posterior} \end{array}$$

The Characterisation Result

The updating satisfies (uninformativeness, symmetry, non-dogmatism, divisibility), iff there exists a homeomorphism $F : \Delta(\Theta) \to \Delta(\Theta)$ such that the updating satisfies

$$
\begin{array}{ccc}
\text{Beliefs} & & \text{Updated Beliefs} \\
\downarrow F & & \uparrow F^{-1} \\
\text{Shadow Prior} & \xrightarrow{\text{Bayes Updating}} & \text{Shadow Posterior}
\end{array}
$$

Equivalently

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

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

Or odds ratio:

$$\frac{F_\theta(u)}{F_{\theta'}(u)} = \frac{F_\theta(\mu)}{F_{\theta'}(\mu)}\frac{p_s^\theta}{p_s^{\theta'}}.$$

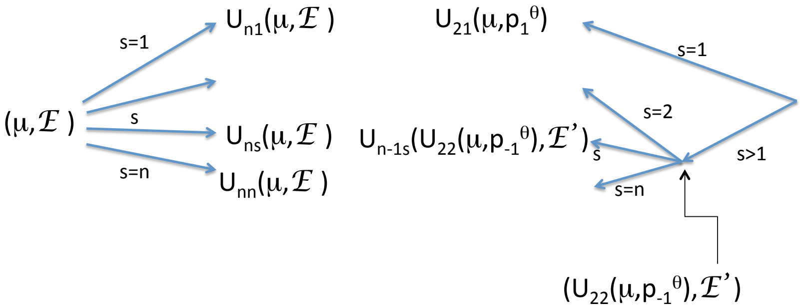
Proof of this Result 1: Simplifying the updating function.

Divisibility and symmetry implies updating has the form

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

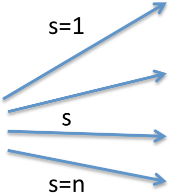
where: $p_s := (p_s^\theta : \theta \in \Theta)$, and $u : \Delta(\Theta) \times [0,1]^{|\Theta|} \to \Delta(\Theta)$. To see this recall

$s = 1$-update depends on only $(p_s^\theta)_{\theta \in \Theta}$



Symmetry implies this is true for any *s*

$(\mu,\mathcal{E})$



$U_{n1}(\mu,\mathcal{E})$ $\quad U_{21}(\mu,p_1)=u(\mu,p_1)$

$U_{ns}(\mu,\mathcal{E})$ $\quad U_{21}(\mu,p_s)=u(\mu,p_s)$

$U_{nn}(\mu,\mathcal{E})$ $\quad U_{21}(\mu,p_n)=u(\mu,p_n)$

$$p_s:=(p_s^{\theta}:\theta\ \ \Theta)$$

u is homogeneous degree zero in p_s

1. Suppose signal 1 is uninformative and consider signal s'
2. Divisibility says
$$U_n = (u(\mu, p_s))_{s \in S}$$
Equals
$$\left[u(\mu, p_1), U_{n-1}\left(u(\mu, \mathbf{1} - p_1), \left(\frac{p_{-1}}{1 - p_1} \right) \right) \right].$$

u is homogeneous degree zero in p_s.

1. If signal 1 is uninformative

$$\left[u(\mu, p_1), U_{n-1}\left(\underbrace{u(\mu, \mathbf{1} - p_1)}_{=\mu}, \left(\frac{p_{-1}}{1 - p_1^{\theta}} \right) \right) \right].$$

2. For signals $s > 1$ we get

$$u(\mu, p_s) \equiv u\left(\mu, \left(\frac{p_s}{1 - p_1} \right) \right)$$

Deriving a Functional Equation

1. If we now re-write the divisibility

$$u(\mu, p_s) \equiv u(u(\mu, \mathbf{1} - p_1), p_s \div (\mathbf{1} - p_1))$$

where $p_s \div (\mathbf{1} - p_1) := (\frac{p_s^\theta}{1 - p_1^\theta})_{\theta \in \Theta}$

2. Hence $u : \Delta^o(\Theta) \times \Delta^o(\Theta) \to \Delta^o(\Theta)$ solves the functional equation

$$u(\mu, \pi) \equiv u(u(\mu, \phi), \pi \div \phi)$$

For all $\mu, \pi, \phi \in \Delta^o(\Theta)$—using homogeneity.

Deriving a Functional Equation

1. If we now re-write the divisibility

$$u(\mu, p_s) \equiv u(u(\mu, \mathbf{1} - p_1), p_s \div (\mathbf{1} - p_1))$$

where $p_s \div (\mathbf{1} - p_1) := (\frac{p_s^\theta}{1 - p_1^\theta})_{\theta \in \Theta}$

2. Hence $u : \Delta^o(\Theta) \times \Delta^o(\Theta) \to \Delta^o(\Theta)$ solves the functional equation

$$u(\mu, \pi) \equiv u(u(\mu, \phi), \pi \div \phi)$$

For all $\mu, \pi, \phi \in \Delta^o(\Theta)$—using homogeneity.

Reducing Dimension

1. Let $w : \Delta^o(\Theta) \to \mathbb{R}_{++}^{|\Theta|-1}$ be

$$w(\mu_1, \ldots, \mu_K) := \left(\frac{\mu_1}{\mu_K}, \ldots, \frac{\mu_{K-1}}{\mu_K}\right).$$

2. Define $\tilde{\mu} := \ln w(\mu) \in \mathbb{R}^{|\Theta|-1}$ and $\tilde{u}, \tilde{\phi}$ and $\tilde{\pi}$ similarly

$\Rightarrow$ transformed functional equation for $\tilde{\mu} : \mathbb{R}^{|\Theta|-1} \times \mathbb{R}^{|\Theta|-1} \to \mathbb{R}^{|\Theta|-1}$

$$\tilde{u}(\tilde{\mu}, \tilde{\pi}) \equiv \tilde{u}\left(\tilde{u}(\tilde{\mu}, \tilde{\phi}), \tilde{\pi} - \tilde{\phi}\right), \qquad \forall \tilde{\mu}, \tilde{\pi}, \tilde{\phi} \in \mathbb{R}^{|\Theta|-1}.$$

Translation Equation

$$\tilde{u}(\tilde{\mu}, x + y) \equiv \tilde{u}(\tilde{u}(\tilde{\mu}, x), y), \qquad \forall \tilde{\mu}, x, y \in \mathbb{R}^{|\Theta|-1}.$$

A simple solution to this multivariate equation is $u(\tilde{\mu}, x) = \tilde{\mu} + x$. This gives Bayesian updating when all the above transformations are reversed.

Translation Equation

$$\tilde{u}(\tilde{\mu}, x + y) \equiv \tilde{u}(\tilde{u}(\tilde{\mu}, x), y), \qquad \forall \tilde{\mu}, x, y \in \mathbb{R}^{|\Theta|-1}.$$

A simple solution to this multivariate equation is $u(\tilde{\mu}, x) = \tilde{\mu} + x$. This gives Bayesian updating when all the above transformations are reversed.

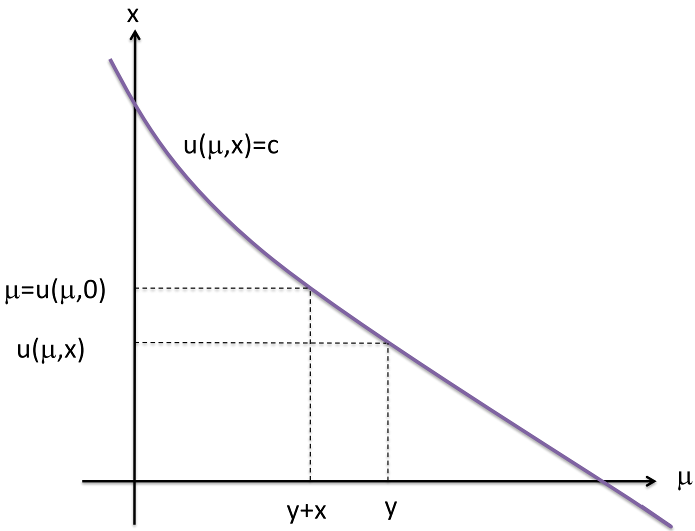
Translation Equation

$$\tilde{u}(\tilde{\mu}, x+y) \equiv \tilde{u}(\tilde{u}(\tilde{\mu}, x), y), \qquad \forall \tilde{\mu}, x, y \in \mathbb{R}^{|\Theta|-1}.$$

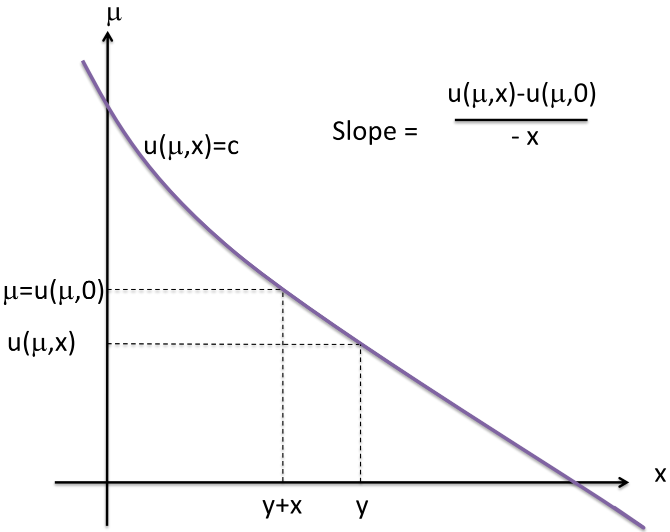
There is a big literature on the classes of solutions to this equation: Aczél and Hosszú (1956), Moszner (1995), Aczél and Dhombres (1989).

- Equation says that $(\tilde{\mu}, x+y)$ and $(\tilde{u}(\tilde{\mu}, x), y)$ are both on the same contour of the $u(.,.)$ function.
- Note that $\tilde{\mu} \equiv \tilde{u}(\tilde{\mu}, 0)$.

Points on a contour

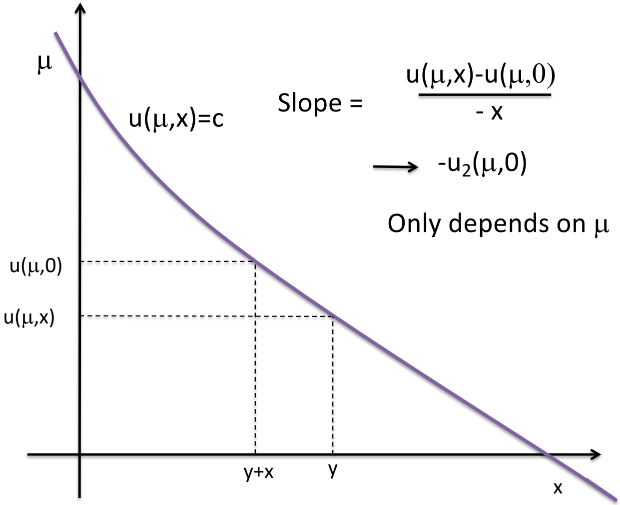


Slope



$$\text{Slope} = \frac{u(\mu,x)-u(\mu,0)}{-x}$$

Slope independent of x



Equation of contours

This implies that all contours have the equation $c = f(\mu) + x$. (Where $f(.)$ is a homeomorphism.)

Thus as the value on the contours is arbitrary we can deduce $u(\mu, x) = g(f(\mu) + x)$ where g is another homeomorphism.

But we know $u(\mu, 0) \equiv \mu$, so $g = f^{-1}$.

Hence all continuous solutions to the functional equation $\tilde{u}(\tilde{\mu}, x + y) \equiv \tilde{u}(\tilde{u}(\tilde{\mu}, x), y)$ have the form

$$\tilde{u}(\mu, x) \equiv f^{-1}(x + f(\mu))$$

where f is a homeomorphism.

The formal proof of Aczél and Hosszú (1956) uses non-dogmatic axiom.

Equation of contours

This implies that all contours have the equation $c = f(\mu) + x$. (Where $f(.)$ is a homeomorphism.)

Thus as the value on the contours is arbitrary we can deduce $u(\mu, x) = g(f(\mu) + x)$ where g is another homeomorphism.

But we know $u(\mu, 0) \equiv \mu$, so $g = f^{-1}$.

Hence all continuous solutions to the functional equation $\tilde{u}(\tilde{\mu}, x + y) \equiv \tilde{u}(\tilde{u}(\tilde{\mu}, x), y)$ have the form

$$\tilde{u}(\mu, x) \equiv f^{-1}(x + f(\mu))$$

where f is a homeomorphism.

The formal proof of Aczél and Hosszú (1956) uses non-dogmatic axiom.

Equation of contours

This implies that all contours have the equation $c = f(\mu) + x$. (Where $f(.)$ is a homeomorphism.)

Thus as the value on the contours is arbitrary we can deduce $u(\mu, x) = g(f(\mu) + x)$ where g is another homeomorphism.

But we know $u(\mu, 0) \equiv \mu$, so $g = f^{-1}$.

Hence all continuous solutions to the functional equation $\tilde{u}(\tilde{\mu}, x + y) \equiv \tilde{u}(\tilde{u}(\tilde{\mu}, x), y)$ have the form

$$\tilde{u}(\mu, x) \equiv f^{-1}(x + f(\mu))$$

where f is a homeomorphism.

The formal proof of Aczél and Hosszú (1956) uses non-dogmatic axiom.

Equation of contours

This implies that all contours have the equation $c = f(\mu) + x$. (Where $f(.)$ is a homeomorphism.)

Thus as the value on the contours is arbitrary we can deduce $u(\mu, x) = g(f(\mu) + x)$ where g is another homeomorphism.

But we know $u(\mu, 0) \equiv \mu$, so $g = f^{-1}$.

Hence all continuous solutions to the functional equation $\tilde{u}(\tilde{\mu}, x + y) \equiv \tilde{u}(\tilde{u}(\tilde{\mu}, x), y)$ have the form

$$\tilde{u}(\mu, x) \equiv f^{-1}(x + f(\mu))$$

where f is a homeomorphism.

The formal proof of Aczél and Hosszú (1956) uses non-dogmatic axiom.

Inverting all the transformations.

- This gives

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

- F is defined so that $e^{f(\ln x)} \circ w \equiv w \circ F$.

Examples of Divisible Non-Bayesian: F

$$F(\mu) = \left(\frac{\mu_1^\alpha}{\sum_\theta \mu_\theta^\alpha}, \ldots, \frac{\mu_K^\alpha}{\sum_\theta \mu_\theta^\alpha} \right)$$

Gives

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

Weighted Bayes, Angrisani, Guarino, Jehiel, and Kitagawa (2017), Bohren and Hauser (2017)

Examples of Divisible Non-Bayesian: F

$$F(\mu) = \left(\frac{\mu_1^\alpha}{\sum_\theta \mu_\theta^\alpha}, \ldots, \frac{\mu_K^\alpha}{\sum_\theta \mu_\theta^\alpha} \right)$$

Gives

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

Weighted Bayes, Angrisani, Guarino, Jehiel, and Kitagawa (2017), Bohren and Hauser (2017)

Examples of Divisible Non-Bayesian F

$$F(\mu) = \left(\frac{e^{-\beta_1/\mu_1}}{\sum_\theta e^{-\beta_\theta/\mu_\theta}}, \ldots, \frac{e^{-\beta_K/\mu_K}}{\sum_\theta e^{-\beta_\theta/\mu_\theta}} \right)$$

Gives

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

"Inverse multinomial logit"

Examples of Divisible Non-Bayesian F

$$F(\mu) = \left(\frac{e^{-\beta_1/\mu_1}}{\sum_\theta e^{-\beta_\theta/\mu_\theta}}, \dots, \frac{e^{-\beta_K/\mu_K}}{\sum_\theta e^{-\beta_\theta/\mu_\theta}} \right)$$

Gives

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

"Inverse multinomial logit"

Relaxing Some Implicit and Explicit Assumptions

- Do not assume 1:1 and dogmatism. Instead suppose the function $\tilde{u}(\mu, x)$ is C1.

$\Rightarrow$ For almost all $\tilde{\mu}$ (excluding a nowhere dense set) the equation $\tilde{u}(\tilde{\mu}, x + y) \equiv \tilde{u}(\tilde{u}(\tilde{\mu}, x), y)$ has a solution of the form

$$\tilde{u}(\mu, x) \equiv f^{-1}(x + f(\mu))$$

on a neighbourhood of $(\mu, 0)$.

Relaxing Some Implicit and Explicit Assumptions

Can allow beliefs to lie in a subspace of $\Delta(\Theta)$, (so the dimension of the set of posteriors is smaller than the dimension of the set of parameters) and have solutions of the form

$$\tilde{u}(\mu, x) \equiv f^{-1}(Cx + f(\mu))$$

where C is an arbitrary matrix of the appropriate dimension that contains a square regular matrix. This admits the same kind of interpretation.

Properties: Consistency?

Consistency: = updating eventually learns/converges to the truth.

Bayes' updating satisfies consistency when parameter spaces are finite or Polish.

$\Rightarrow$ Divisible updating is consistent (provided you don't choose a silly F).

For all θ there exists $\mu^\infty \in \Delta(\Theta)$ such that $\mu^t \to \mu^\infty$, $\mathbb{P}^\theta$ almost surely.

If $U_n(e_\theta, \mathcal{E}) = (e_\theta, \ldots, e_\theta)$ $p^\theta \neq p^{\theta'}$ for all $\theta' \neq \theta$, and $\mu^0 \in \Delta^o(\Theta)$, then, $\mu^\infty = e_\theta$ with $\mathbb{P}^\theta$ probability one.

Properties: Consistency?

Consistency: = updating eventually learns/converges to the truth.

Bayes' updating satisfies consistency when parameter spaces are finite or Polish.

$\Rightarrow$ Divisible updating is consistent (provided you don't choose a silly F).

For all θ there exists $\mu^\infty \in \Delta(\Theta)$ such that $\mu^t \to \mu^\infty$, $\mathbb{P}^\theta$ almost surely.

If $U_n(e_\theta, \mathcal{E}) = (e_\theta, \ldots, e_\theta)$ $p^\theta \neq p^{\theta'}$ for all $\theta' \neq \theta$, and $\mu^0 \in \Delta^o(\Theta)$, then, $\mu^\infty = e_\theta$ with $\mathbb{P}^\theta$ probability one.

Properties: Consistency?

Consistency: = updating eventually learns/converges to the truth.

Bayes' updating satisfies consistency when parameter spaces are finite or Polish.

$\Rightarrow$ Divisible updating is consistent (provided you don't choose a silly F).

For all θ there exists $\mu^{\infty} \in \Delta(\Theta)$ such that $\mu^{t} \rightarrow \mu^{\infty}$, $\mathbb{P}^{\theta}$ almost surely.

If $U_n(e_{\theta}, \mathcal{E}) = (e_{\theta}, \ldots, e_{\theta})$ $p^{\theta} \neq p^{\theta'}$ for all $\theta' \neq \theta$, and $\mu^{0} \in \Delta^{o}(\Theta)$, then, $\mu^{\infty} = e_{\theta}$ with $\mathbb{P}^{\theta}$ probability one.

Biases in the Learning?

Bayes' updating $\Rightarrow$ belief in the parameter θ on average increases when θ is true (Submartingale).

The convexity of the homeomorphism is what matters here:

Divisible updating is

$$\begin{aligned} \text{Locally consistent} \quad &\Leftrightarrow \quad \mu_\theta \leq E^\theta(u_\theta(\mu, p_s)), \\ \text{Locally inconsistent} \quad &\Leftrightarrow \quad \mu_\theta > E^\theta(u_\theta(\mu, p_s)). \end{aligned}$$

Biases in the Learning?

Bayes' updating $\Rightarrow$ belief in the parameter θ on average increases when θ is true (Submartingale).

The convexity of the homeomorphism is what matters here:

Divisible updating is

$$\text{Locally consistent} \quad \Leftrightarrow \quad \mu_\theta \leq E^\theta(u_\theta(\mu, p_s)),$$

$$\text{Locally inconsistent} \quad \Leftrightarrow \quad \mu_\theta > E^\theta(u_\theta(\mu, p_s)).$$

Biases in the Learning?

Bayes' updating $\Rightarrow$ belief in the parameter θ on average increases when θ is true (Submartingale).

The convexity of the homeomorphism is what matters here:

Divisible updating is

$$\begin{aligned} &\text{Locally consistent} && \Leftrightarrow && \mu_\theta \leq E^\theta(u_\theta(\mu, p_s)), \\ &\text{Locally inconsistent} && \Leftrightarrow && \mu_\theta > E^\theta(u_\theta(\mu, p_s)). \end{aligned}$$

Biases in the Learning

The Bayes' updating after the homeomorphism has been applied has a likelihood ratio that is a conditional martingale

$$E^{\theta}\left(\frac{1-f(u_\theta)}{f(u_\theta)}\right) = \frac{1-f(\mu_\theta)}{f(\mu_\theta)}$$

Applying Jensen's and the monotonicity of $f(.) \Rightarrow$

$$\mu_\theta \leq E^{\theta}(u_\theta(\mu, p_s)) \qquad \text{if } \frac{1}{f(.)} \text{ is convex.}$$

$$\mu_\theta \geq E^{\theta}(u_\theta(\mu, p_s)) \qquad \text{if } \frac{1}{f(.)} \text{ is concave.}$$

Biases in the Learning

The Bayes' updating after the homeomorphism has been applied has a likelihood ratio that is a conditional martingale

$$E^{\theta}\left(\frac{1-f(u_{\theta})}{f(u_{\theta})}\right)=\frac{1-f(\mu_{\theta})}{f(\mu_{\theta})}$$

Applying Jensen's and the monotonicity of $f(.)$ $\Rightarrow$

$$\mu_{\theta}\leq E^{\theta}(u_{\theta}(\mu,p_s)) \qquad \text{if } \frac{1}{f(.)} \text{ is convex.}$$

$$\mu_{\theta}\geq E^{\theta}(u_{\theta}(\mu,p_s)) \qquad \text{if } \frac{1}{f(.)} \text{ is concave.}$$

Under/Over-reaction?

Bayes' updating

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

How does the presence of a map F affect this variance? There are two issues

1. If F^{-1} moves points further apart it exaggerates the variability of Bayes. (Slope of F.)
2. If F maps points to extremities then little updating.

Under/Over-reaction?

Bayes' updating

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

How does the presence of a map F affect this variance? There are two issues

1. If F^{-1} moves points further apart it exaggerates the variability of Bayes. (Slope of F.)
2. If F maps points to extremities then little updating.

Under/Over-reaction?

Overreaction result:

$$\text{Var}\left[\log \frac{u_\theta(\mu, p_s)}{1 - u_\theta(\mu, p_s)}\right] \geq \text{Var}\left[\log \frac{p^\theta}{p^{\theta'}}\right].$$

If $f'(\mu) < f(\mu)(1 - f(\mu))/(\mu(1 - \mu))$ for all μ.

Underreaction result:

$$\text{Var}\left[\log \frac{u_\theta(\mu, p_s)}{1 - u_\theta(\mu, p_s)}\right] \leq \text{Var}\left[\log \frac{p^\theta}{p^{\theta'}}\right].$$

If $f'(\mu) > f(\mu)(1 - f(\mu))/(\mu(1 - \mu))$ for all μ.

Under/Over-reaction?

Overreaction result:

$$\text{Var}\left[\log \frac{u_\theta(\mu, p_s)}{1 - u_\theta(\mu, p_s)}\right] \geq \text{Var}\left[\log \frac{p^\theta}{p^{\theta'}}\right].$$

If $f'(\mu) < f(\mu)(1 - f(\mu))/(\mu(1 - \mu))$ for all μ.

Underreaction result:

$$\text{Var}\left[\log \frac{u_\theta(\mu, p_s)}{1 - u_\theta(\mu, p_s)}\right] \leq \text{Var}\left[\log \frac{p^\theta}{p^{\theta'}}\right].$$

If $f'(\mu) > f(\mu)(1 - f(\mu))/(\mu(1 - \mu))$ for all μ.

Unbiased/Bayes' Plausible/Martingale Updating

This is the property that the expected value of the posterior beliefs equals the prior beliefs. For any $\mu > 0$, $n > 1$, and $\mathcal{E} \in \Delta^o(S)^K$

$$\mu \equiv \sum_{s \in S} \left(\sum_{\theta \in \Theta} \mu_\theta p_s^\theta \right) U_n^s(\mu, (p)_{\theta \in \Theta}).$$

- Difficult to explain to subjects and motivate normatively.
- Characterisation Result: the updating function $U_n(\mu, \mathcal{E})$ is unbiased if and only if it is the Bayesian update for some misspecified experiment $\mathcal{E}'$.

Sufficient Conditions for Full Bayes

Result: Bayesian updating is the only updating that satisfies: Uninformativeness, Symmetry, Divisibility, Non-dogmatic, and Unbiasedness.

Why?

Suppose you have a binary experiment that either reveals the state θ if it is true but is otherwise uninformative, then

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

(where e_θ is a vector with one in the θth entry and zeros elsewhere and y_θ has zero in the θth entry. Hence

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

So $1 = F_\theta^{-1}(e_\theta)$

Sufficient Conditions for Full Bayes

Result: Bayesian updating is the only updating that satisfies: Uninformativeness, Symmetry, Divisibility, Non-dogmatic, and Unbiasedness.
Suppose the binary experiment reveals the state θ with probability p^θ if it is true, then

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

or

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

So $\mu_\theta = F_\theta(\mu)$

What's missing?

- Domain and range of the function
- Discrete Domain
- Random updates
- Local updates

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.

AHN, D. S., F. ECHENIQUE, AND K. SAITO (2018): "On path independent stochastic choice," *Theoretical Economics*, 13(1), 61–85.

ANGRISANI, M., A. GUARINO, P. JEHIEL, AND T. KITAGAWA (2017): "Information Redundancy Neglect Versus Overconfidence: A Social Learning Experiment," Cemmap working paper, UCL.

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.

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.

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.

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