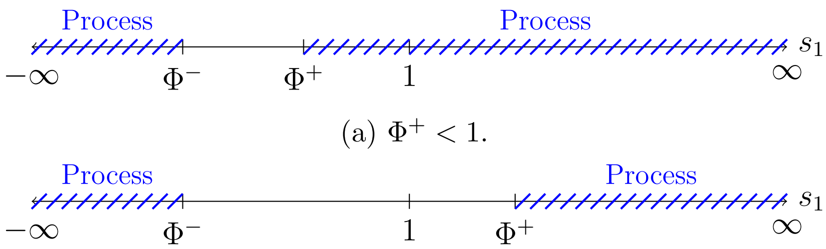



(a) $\Phi^+ < 1$.

(b) $\Phi^+ \geq 1$

Figure 2: Equilibrium Information Processing Strategy: the DM processes $s_1$ if and only if its realization falls on the shaded area.

*tions:*

$$\left.\begin{aligned} \Phi^+ &= \frac{u_R\, p_R}{u_L\, p_L} STR\left(s_2 \in \left(0, \frac{u_R\, p_R}{u_L\, p_L \dfrac{F_{1L}(\Phi^+) - F_{1L}(1/\Phi^-)}{F_{1R}(\Phi^+) - F_{1R}(1/\Phi^-)}}\right)\right) \\ \Phi^- &= \frac{u_L\, p_L}{u_R\, p_R} STR\left(s_2 \in \left[\frac{u_R\, p_R}{u_L\, p_L \dfrac{F_{1L}(\Phi^+) - F_{1L}(1/\Phi^-)}{F_{1R}(\Phi^+) - F_{1R}(1/\Phi^-)}}, \infty\right)\right) \end{aligned}\right\} \tag{10}$$

The proof is very similar to that of proposition 1 and is therefore skipped. Different from the unaware case, the DM may process weak belief-challenging information (when $\Phi^+ < 1$, as shown in figure 2). This difference is driven by the self-fulfilling nature of the game: suppose the DM's period 2 self conjectures that his period 1 self ignores strong belief-challenging signals but processes weak belief-challenging signals. In the view of his period 1 self, he knows that if he ignores the weak belief-challenging signals, his period 2 self will mistakenly infer that $s_1$ strongly supports state $R$. It in turn gives the DM's period 1 self incentive to process weak belief-challenging signals and take action $l$, as otherwise he will be under-confident about state $L$ and switch to action $r$ too easily in period 2.

Despite the slight difference, the behavior implications hold qualitatively. First, the DM processes strong enough signals, but may ignore weak signals, *i.e.*, it resembles a preference for strong signals (although it is weaker than the version in the unaware case). Moreover, by equation (10), the limits of $\Phi^+$ and $\Phi^-$ when $u_L p_L / u_R p_R \to +\infty$ are the same as in the unaware case[32]. Hence, when $u_L p_L / u_R p_R$ is large enough, $\Phi^- < \Phi^+$. If the DM processes a signal $s^{-1}$ that supports state $R$, he also processes a equally strong signal $s$ that supports state $L$. The reverse is not necessarily true.

**Proposition 7.** *The equilibrium processing strategy of the DM explains the following behavioral phenomena:*

1. ***(Preference for strong signals)***
   *The DM processes a signal $s_1$ if it is strong enough.*

2. ***(Confirmation bias for confident individuals)***
   *If the DM a priori believes strongly that state L is true, his processing strategy exhibits confirmation bias, i.e., $\Phi^- > \Phi^+$ if $p_L$ is large enough.*

3. ***Wishful thinking if one state is much more desirable than the others***
   *If state L is much more desirable than state R, the DM's processing strategy exhibits confirmation bias, i.e., $\Phi^- > \Phi^+$ if $u_L/u_R$ is large enough.*

Note that the results hold in all equilibria in the aware case, *i.e.,* all equilibrium processing strategies of the DM have (qualitatively) the same behavioral implications as in the unaware case.

Lastly, because there is an issue of multiple equilibria, the result of the comparative analysis would not be as clean as in the unaware case. However, the insights still hold true. As shown in equation (10), the thresholds which characterize the equilibrium processing strategy depend only on a subset of all realizations of $s_2$. As in the unaware case, providing in average more informative signal to the DM does not necessarily increases the average strength

[32]That is, as shown in the proof of proposition 2,

$$\lim_{u_L p_L / u_R p_R \to +\infty} \Phi^+ = 1$$
$$\lim_{u_L p_L / u_R p_R \to +\infty} \Phi^- = +\infty$$

of a subset of all realizations of $s_2$. Therefore, it could strengthen the confirmation bias of the DM.

# 8 Applications

In this section, I provide two simple examples, which relates to polarization and media competition in the presence of information overload. For simplicity, I only look at the case where $u_L = u_R = 1$, $T = 2$ and $\bar{T} = 1$. Moreover, as in the illustrative example, I assume $s_1$ and $s_2$ follow a symmetric information structure with three possible realizations, denoted as $1/q, 1, q$ where $q > 1$. The information structure is represented in the following table:

| | $f_{t\omega}(1/q)$ | $f_{t\omega}(1)$ | $f_{t\omega}(q)$ |
|---|---|---|---|
| $\omega = L$ | $\lambda(1+q)^{-1}$ | $1-\lambda$ | $\lambda q(1+q)^{-1}$ |
| $\omega = R$ | $\lambda q(1+q)^{-1}$ | $1-\lambda$ | $\lambda(1+q)^{-1}$ |

Table 6: The distribution of $s_t$, $t = 1, 2$, given the state of the world.

An increase in $\lambda$ represents a better access to valuable information while $q$ represents the strength/quality of the informative signals. In the illustrative example, $\lambda = 0.3$ and $q = 2.5$. Throughout the section, I focus on the unaware case and assume $g$ is degenerate.

## 8.1 Polarization

Many empirical studies have documented the phenomenon of political polarization or stronger partisanship in the US in recent years. Bartels (1998) and Bartels (2000) show that party identification has become a better predictor of vote decisions and document a decline in volatility of election outcomes. Moreover, the polarization is stronger among citizens who are more politically engaged and partisan. (See Evans (2003), Baldassarri and Gelman (2008), Abramowitz and Saunders (2008) and Hetherington (2009)). In contrast, Gentzkow and Shapiro (2011) and Flaxman et al. (2016) show that online media exposes individuals to belief-challenging information, although somewhat counter-intuitively, Flaxman et al. (2016) also find that online media are associated to an increase in ideological distance between individuals.

In the following, I show how the model of limited cognitive ability presented in this paper sheds light on the wide range of phenomena documented in the literature. Assume that Alice and Bob have opposite prior beliefs about the state of the world with obvious notations $p_L^a = p_R^b = p > 1/2$. The informative signals are strong enough, *i.e.*, $q > \frac{p}{1-p}$, which rules out the trivial case where the two individuals always take their a priori optimal action. Without loss of generality, I assume that $L$ is the true state.

I analyze three indicators of polarization:

1. the probability that both individuals take the same action, which is denoted as $P_{concensus}$. It is used to asses the intuition where more/better information results in a higher probability of achieving consensus;

2. the probability that the individual takes his/her a priori optimal action, which is denoted as $P^j_{default}$, $j = A, B$. It measures how well the prior belief, or analogously party identification, predicts voting decision, which is studied in the empirical literature of political science;

3. the change in the distance between the beliefs of the two individuals after receiving information. It corresponds to belief polarization and is widely analyzed in the theoretical economics literature[33].

Proposition 1 allows me to characterize the processing strategy of the two individuals as follows:

**Corollary 7.** *In period 1, both Alice and Bob process $s_1$ if it is belief-confirming and ignore it if it is pure noise. There exists some thresholds $q^-$, $\lambda^-$ and $p^-$ such that they process belief-challenging signal if and only if*

- *their prior is weak enough, i.e., $p < p^-$, or;*

[33] For example, see Acemoglu et al. (2007), Baldassarri and Gelman (2008).

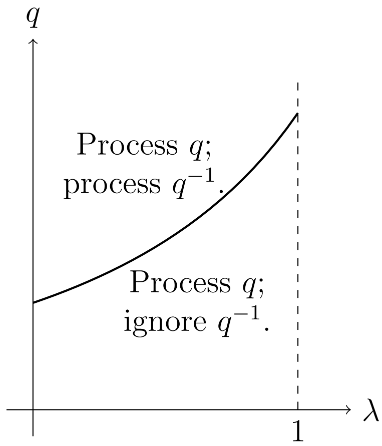



Figure 3: Alice's processing strategy in period 1 as a function of $\lambda$ and $q$, fixing $p_L^a = p = 0.7$.

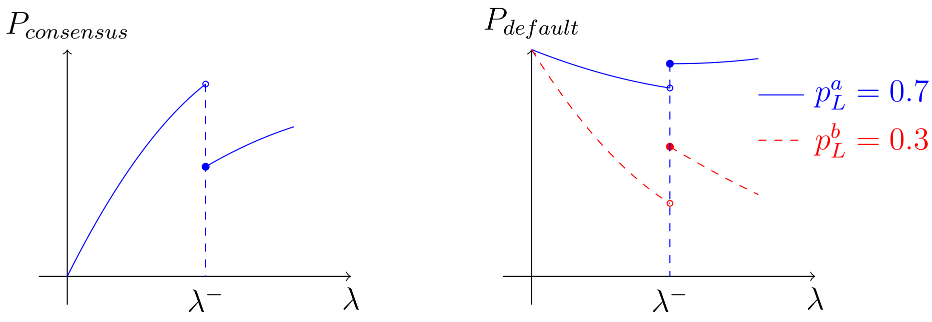
- *the informative signals are strong enough, i.e.,* $q > q^-$ *, or;*
- *the access to information is poor enough, i.e.,* $\lambda < \lambda^-$.

The proof is shown in the appendix and the result is illustrated in figure 3. The following proposition, illustrated in figure 4, presents how a change in the access to information $\lambda$ affects the first two indicators of polarization, $P_{consensus}$ and $P^j_{default}$.

**Proposition 8.** *The probability that Alice and Bob take the same action,* $P_{consensus}$*, is non-monotonic in* $\lambda$*, i.e., it is increasing in the range* $[0, \lambda^-)$ *and* $[\lambda^-, 1]$*, but exhibits a downward jump at* $\lambda = \lambda^-$.

*Similarly, the probability that Alice/Bob takes her/his a priori optimal action,* $P^a_{default}$ *or* $P^b_{default}$*, is also non-monotonic in* $\lambda$*, i.e., both are decreasing in* $\lambda$ *in the range* $[0, \lambda^-)$*, but exhibit a upward jump at* $\lambda = \lambda^-$.

When there is better access to information, fixing the processing strategies of the two individuals, the probability that they takes the optimal action $l$ increases. However, as shown in proposition 3 and corollary 7, a better access to information promotes biased processing behavior, *i.e.,* when $\lambda$ is big enough, the two individuals ignore belief-challenging information. As a result, it could reduce the probability of achieving consensus and increase the




(a) $P_{consensus}$ with $p = 0.7$ and $\Phi = 0.7$.

(b) $P_{default}$ with $\Phi = 0.7$

Figure 4: The effect of $\lambda$ on the first two indicators of polarization.

probability that the individuals take their a priori optimal action, despite the availability of more valuable public information. The limitation in processing ability hinders the benefits of information technology because individuals strategically allocate their cognitive resources in the presence of information overload.

Moreover, as individuals with different prior beliefs adopt different processing strategies, their beliefs can be polarized even if they receive the same sequence of information.

**Corollary 8.** *When $\lambda \geq \lambda^-$, $q \leq q^-$ or $p \geq p^-$, the distance between the beliefs of Alice and Bob increases after receiving information if the signals in the two periods support different states. More specifically, if $s_1 > 1 > s_2$ or $s_2 > 1 > s_1$, Alice becomes more confident that state $L$ is true while Bob becomes more confident that state $R$ is true.*

*Moreover, the probability that $s_1 > 1 > s_2$ or $s_2 > 1 > s_1$ increases in $\lambda$ and decreases in $q$.*

*Proof.* Without loss of generality, I analyze the case where $s_1 = q$ and $s_2 = 1/q$. When $\lambda \geq \lambda^-$, $q \leq q^-$ or $p \geq p^-$, Alice and Bob process belief-confirming information but ignore belief-challenging information in period 1. Therefore, Alice processes $s_1$ while Bob ignores $s_1$ and processes $s_2$. Their beliefs in period 3 are:

$$\Pr_L^a(\mathscr{M}_3^a) = \Pr_L^a(q) = \left(1 + \frac{1-p}{pq}\right)^{-1} > p_L^a \qquad ;$$

$$\Pr_L^b(\mathscr{M}_3^b) = \Pr_L^b(q^{-1}) = \left(1 + \frac{p}{(1-p)q^{-1}}\right)^{-1} < p_L^b.$$

On the other hand, the probability that $s_1 > 1 > s_2$ or $s_2 > 1 > s_1$ equals

$$2\lambda^2 q(1+q)^{-2},$$

which increases in $\lambda$ and decreases in $q$. □

Corollary 8 shows that even when the same sequence of information is available for both individuals, the difference in their prior beliefs induces different processing strategies and could polarize their beliefs. Moreover, it happens only when there are sufficiently good access to information, or when the prior beliefs of the two individuals are strong enough. In other words, when the two individuals are partisan enough, a better access to information gives rise to the possibility of polarization even under a setting with public information. This result sheds light on the empirical evidence that polarization is much stronger among individuals who are more partisan.

On the other hand, belief polarization happens when the information available in the two periods are contradictory, *i.e.*, $s_1 > 1 > s_2$ or $s_2 > 1 > s_1$. Its probability increases when there is better access to information and when the quality of information decreases. Arguably, the Internet contributes to both. While it provides us with enormous amount of information, it also facilitates the spread of rumors, fake news and low-quality information. It is much easier to find contradictory information on the Internet. It gives more incentive for individuals to selectively attend to belief-confirming information, which increases the probability of belief polarization.

## 8.2 Media Competition

In this second application, I study media strategy in the information era. In order to introduce the role of the media, I present a variation of the main model.

Formally, there is a continuum of media, indexed by $i \in \mathscr{I}$. In period $t = 1, 2$, each media $i$ collects a signal $s_{it}$ about the state of the world and publishes a piece of news $m_{it}$. The signal $s_{it}$ collected by different media are independent and follow the distribution defined in the beginning of the section, *i.e.,* table 6. I assume that the media cannot post fake news and therefore either publishes the signal it receives or publishes nothing, *i.e.,* $m_{it} \in \{\emptyset, s_{it}\}$. To simplify the analysis, I assume that there are three types of non-strategic (biased) media, $\{T_L, T_R, T_N\}$, which publish according to the following fixed rules:

- media $\{i \mid i \in T_L\}$ is biased towards state $L$: he publishes $s_{it}$ if it supports state $L$ ($s_{it} = q$), but publishes $\emptyset$ if he receives $s_{it} \in \{1/q, 1\}$;
- media $\{i \mid i \in T_R\}$ has an opposite bias: he publishes a signal if it supports state $R$ ($s_{it} = 1/q$) but publishes $\emptyset$ if he receives $s_i \in \{1, q\}$;
- media $\{i \mid i \in T_N\}$ has no bias and publishes any informative signal he receives, *i.e.,* $m_{it} = s_i$ if and only if $s_i \neq 1$.

I assume that each media belongs to one and only one of the three types, and the type of each media is a public information. The time line of the game is as follows:

**Period 1** The DM chooses which media he visits and processes the piece of news $m_{i1}$ posted by the media.

**Period 2** If the media visited by the DM in period 1 published nothing, he chooses again a media outlet to visit. Otherwise, he visits no media as processing the information in period 1 takes time.

**Period 3** The DM forms his belief with his memory of information and takes an action $l$ or $r$.

Note that this variation differs from the main model only in terms of interpretation. Here the DM control his "diet" of information by choosing which (biased or unbiased) media to visit, instead of choosing whether to process or ignore the signals he receives. For example, if the DM visits media $i \in T_L$ in period 1, it is as if he chooses to process the period 1 signal if and only if it supports state $L$. By corollary 7, the DM visits the biased media in period 1 if his prior belief is strong enough.

**Corollary 9.** *Suppose the DM has prior belief* $(p_L, 1-p_L)$ *where* $p_L \geq 1/2$. *In period 1, there exists a threshold* $p^- \in (1/2, 1)$ *such that if* $p_L \geq p^-$, *he visits a media* $i$ *where* $i \in T_L$*; otherwise, he visits a media* $i$ *where* $i \in T_N$.

*If the media visited by the DM in period 1 published nothing, he visits a media* $i$ *where* $i \in T_N$ *in period 2.*

Now I turn to analyze the viewership of the three types of media. In the society, individuals' prior belief $p_L$ are distributed according to $g(p)$ where $g(p) > 0$ for all $p \in (0,1)$. Its c.d.f. is denoted by $G$. Define the viewership of media $\{i \mid i \in T_j\}$ as the mass of individuals which visit media $\{i \mid i \in T_j\}$ across the two periods. Denote it by $V_j$ for $j = L, R, N$. By corollary 9,

$$
\begin{aligned}
V_L &= 1 - G(p^-); \\
V_R &= G(1 - p^-); \\
V_N &= 2 - V_L - V_R - \lambda.
\end{aligned}
$$

That is, biased media attracts views from individuals with strong beliefs, while unbiased media serves the others.

**Corollary 10.** *When there are better access to information* ($\lambda$ *increases), the viewership of the biased media increases while the viewership of unbiased media decreases, i.e.,* $V_L$ *and* $V_R$ *increase with* $\lambda$ *while* $V_N$ *decreases in* $\lambda$.

*On the other hand, when the quality of information increases (*$q$ *increases), the viewership of the biased media decreases while the viewership of unbiased media increases, i.e.* $V_L$ *and* $V_R$ *decrease with* $q$ *while* $V_N$ *increases in* $q$.

*Proof.* The result follows from corollary 7. □

As shown in previous sections, a better access to information or an decrease in quality of information strengthens the confirmation bias of individuals, which in turn increases the viewership of the biased media. The increase in viewership increases the profitability of biased media and thus incentivize media to adopt a biased strategy. This result sheds light on the emergence of partisan media in recent years as the Internet promotes biased processing behavior.

# 9 Conclusion

In conclusion, this paper investigates the information processing behavior of a decision maker who can process only a subset of all available signals. I show that this limitation in processing ability drives a number of well-documented behavioral "biases", including preference of strong signals, confirmation bias for confident individuals and wishful thinking.

These "biases" has been attracting lots of attention in the behavioral economics literature, in which many have analyzed how these "biases" affects different market outcomes by introducing directly the "biases" in traditional economics models. In contrast, instead of taking the "biases" as they are, this paper aims to improve our understanding by analyzing their cause. In particular, I show that these "biases" are features of optimal strategies if we take into account our limited cognitive ability as a human being.

This approach allows us to analyze the "biases" as an outcome of an optimization problem. It brings two advantages. First, not only that it explains the existence of the "biases", but using standard techniques of comparative statistics, it also explains how the "biases" change among individuals with different personal characteristics or abilities, and how they change in different situations faced by the individuals. Second, it allows us to study how regulatory policies could play a role in changing these "biases" and associated market outcomes.

Thus, looking forward and building from the insights of this paper, there are two different ways to further develop the literature. First, in policy analysis with behavioral settings, modeling the "biases" as optimal strategies allows us to take into account the indirect effects of policy interventions on behaviors of individuals. It contributes to a more complete analysis than if we take the "biases" as they are. For example (loosely speaking), if providing more information to consumers strengthens their confirming bias, it might back fire as it could weaken competition and increase prices. Second, more experimental or empirical work has to be done to understand how "biases" are formed and vary across different individuals or settings. It will give us a better understanding on whether the "biases" do indeed relate to the limitation in ability. And as an (early) answer to that question, an ongoing experimental study conducted by myself and co-authors, Goette et al. (2018), does find evidences that a larger cognitive load strengthens confirmation bias in belief formation.

# Reference

[1] Alan Abramowitz and Kyle Saunders. "Is polarization a myth?" In: *The Journal of Politics* 70.02 (2008), pp. 542–555.

[2] Daron Acemoglu, Victor Chernozhukov, and Muhamet Yildiz. "Learning and Disagreement in an Uncertain World". In: *Working Paper* (2007).

[3] Philippe Aghion and Jean Tirole. "Formal and real authority in organizations". In: *Journal of Political Economy* 105.1 (1997), pp. 1–29.

[4] George Akerlof and William Dickens. "The economic consequences of cognitive dissonance". In: *American Economic Review* 72.3 (1982), pp. 307–319.

[5] Delia Baldassarri and Andrew Gelman. "Partisans without constraint: Political polarization and trends in American public opinion". In: *American Journal of Sociology* 114.2 (2008), pp. 408–446.

[6] Sandeep Baliga and Jeffrey C Ely. "Mnemonomics: the sunk cost fallacy as a memory kludge". In: *American Economic Journal: Microeconomics* 3.4 (2011), pp. 35–67.

[7] Larry M Bartels. "Electoral continuity and change, 1868–1996". In: *Electoral Studies* 17.3 (1998), pp. 301–326.

[8] Larry M Bartels. "Partisanship and voting behavior, 1952-1996". In: *American Journal of Political Science* (2000), pp. 35–50.

[9] Roland Bénabou and Jean Tirole. "Self-confidence and personal motivation". In: *Quarterly Journal of Economics* (2002), pp. 871–915.

[10] Aaron Benjamin and Robert Bjork. "On the relationship between recognition speed and accuracy for words rehearsed via rote versus elaborative rehearsal". In: *Journal of Experimental Psychology: Learning, Memory, and Cognition* 26.3 (2000), p. 638.

[11] James B Brewer, Zuo Zhao, John E Desmond, Gary H Glover, and John DE Gabrieli. “Making memories: brain activity that predicts how well visual experience will be remembered”. In: *Science* 281.5380 (1998), pp. 1185–1187.

[12] Markus Brunnermeier and Jonathan Parker. “Optimal expectations”. In: *American Economic Review* 95.4 (2005), pp. 1092–1118.

[13] Juan D Carrillo and Thomas Mariotti. “Strategic ignorance as a self-disciplining device”. In: *Review of Economic Studies* 67.3 (2000), pp. 529–544.

[14] Max Coltheart, Robyn Langdon, and Ryan McKay. “Delusional belief”. In: *Annual Review of Psychology* 62 (2011), pp. 271–298.

[15] Olivier Compte and Andrew Postlewaite. “Belief Formation”. In: *Unpublished* (2012).

[16] Michael H Connors and Peter W Halligan. “A cognitive account of belief: a tentative road map”. In: *Frontiers in Psychology* 5 (2015), p. 1588.

[17] Fergus Craik, Richard Govoni, Moshe Naveh-Benjamin, and Nicole D Anderson. “The effects of divided attention on encoding and retrieval processes in human memory”. In: *Journal of Experimental Psychology: General* 125.2 (1996), p. 159.

[18] Fergus Craik and Robert S Lockhart. “Levels of processing: A framework for memory research”. In: *Journal of Verbal Learning and Verbal Behavior* 11.6 (1972), pp. 671–684.

[19] Fergus Craik and Endel Tulving. “Depth of processing and the retention of words in episodic memory”. In: *Journal of Experimental Psychology: General* 104.3 (1975), p. 268.

[20] Jacques Crémer. “Arm’s length relationships”. In: *The Quarterly Journal of Economics* 110.2 (1995), pp. 275–295.

[21] David Dunning and Emily Balcetis. “Wishful seeing: How preferences shape visual perception”. In: *Current directions in psychological science* 22.1 (2013), pp. 33–37.

[22] David Eil and Justin Rao. "The good news-bad news effect: asymmetric processing of objective information about yourself". In: *American Economic Journal: Microeconomics* 3.2 (2011), pp. 114–138.

[23] Benjamin Enke and Florian Zimmermann. "Correlation neglect in belief formation". In: *Review of Economics Studies* (forthcoming).

[24] John H Evans. "Have Americans' Attitudes Become More Polarized? An Update". In: *Social Science Quarterly* 84.1 (2003), pp. 71–90.

[25] Erik Eyster and Matthew Rabin. "Cursed equilibrium". In: *Econometrica* 73.5 (2005), pp. 1623–1672.

[26] Seth Flaxman, Sharad Goel, and Justin Rao. "Filter bubbles, echo chambers, and online news consumption". In: *Public Opinion Quarterly* (2016), pp. 298–320.

[27] Matthew Gentzkow and Jesse M Shapiro. "Ideological Segregation Online and Offline". In: *Quarterly Journal of Economics* 126.4 (2011), pp. 1799–1839.

[28] Lorenz Goette, Hua-Jing Han, and Benson Tsz Kin Leung. "Memory and (Biased) Belief Formation". In: *Working Paper* (2018).

[29] Daniel Goldstein, Preston McAfee, and Siddharth Suri. "The effects of exposure time on memory of display advertisements". In: *Proceedings of the 12th ACM Conference on Electronic Commerce*. ACM. 2011, pp. 49–58.

[30] Russell Golman, David Hagmann, and George Loewenstein. "Information avoidance". In: *Journal of Economics Literature* 55.1 (2017), pp. 96–135.

[31] Marc Hetherington. "Review article: Putting polarization in perspective". In: *British Journal of Political Science* 39.02 (2009), pp. 413–448.

[32] Laurent Itti and Pierre Baldi. "Bayesian surprise attracts human attention". In: *Advances in Neural Information Processing Systems* 18 (2006), p. 547.

[33] Philippe Jehiel. “Analogy-based expectation equilibrium”. In: *Journal of Economic theory* 123.2 (2005), pp. 81–104.

[34] Philippe Jehiel and Jakub Steiner. “On Second Thoughts, Selective Memory, and Resulting Behavioral Biases”. In: *Unpublished* (2018).

[35] Dan M Kahan, Ellen Peters, Maggie Wittlin, Paul Slovic, Lisa Larrimore Ouellette, Donald Braman, and Gregory Mandel. “The polarizing impact of science literacy and numeracy on perceived climate change risks”. In: *Nature Climate Change* 2.10 (2012), pp. 732–735.

[36] Niklas Karlsson, George Loewenstein, and Duane Seppi. “The ostrich effect: Selective attention to information”. In: *Journal of Risk and Uncertainty* 38.2 (2009), pp. 95–115.

[37] Botond Kőszegi. “Health anxiety and patient behavior”. In: *Journal of Health Economics* 22.6 (2003), pp. 1073–1084.

[38] Zlatan Krizan and Paul D Windschitl. “The influence of outcome desirability on optimism”. In: *Psychological Bulletin* 133.1 (2007), p. 95.

[39] Robyn Langdon and Max Coltheart. “The cognitive neuropsychology of delusions”. In: *Mind & Language* 15.1 (2000), pp. 184–218.

[40] Filip Matejka and Alisdair McKay. “Rational inattention to discrete choices: A new foundation for the multinomial logit model”. In: *American Economic Review* 105.1 (2014), pp. 272–298.

[41] Markus Prior. “Media and political polarization”. In: *Annual Review of Political Science* 16 (2013), pp. 101–127.

[42] T Shallice, P Fletcher, CD Frith, P Grasby, RSJ Frackowiak, and RJ Dolan. “Brain regions associated with acquisition and retrieval of verbal episodic memory”. In: *Nature* 368.6472 (1994), pp. 633–635.

[43] Nachum Sicherman, George Loewenstein, Duane J Seppi, and Stephen P Utkus. “Editor’s Choice Financial Attention”. In: *Review of Financial Studies* 29.4 (2016), pp. 863–897.

[44] Christopher A Sims. “Implications of rational inattention”. In: *Journal of Monetary Economics* 50.3 (2003), pp. 665–690.

[45] Patrick S Sullivan, Amy Lansky, Amy Drake, and HITS-2000 Investigators. "Failure to return for HIV test results among persons at high risk for HIV infection: results from a multistate interview project". In: *JAIDS Journal of Acquired Immune Deficiency Syndromes* 35.5 (2004), pp. 511–518.

[46] Amos Tversky and Daniel Kahneman. "Availability: A heuristic for judging frequency and probability". In: *Cognitive Psychology* 5.2 (1973), pp. 207–232.

[47] Melina Uncapher and Michael Rugg. "Encoding and the durability of episodic memory: a functional magnetic resonance imaging study". In: *The Journal of Neuroscience* 25.31 (2005), pp. 7260–7267.

[48] Anthony Wagner, Daniel Schacter, Michael Rotte, Wilma Koutstaal, Anat Maril, Anders Dale, Bruce Rosen, and Randy Buckner. "Building memories: remembering and forgetting of verbal experiences as predicted by brain activity". In: *Science* 281.5380 (1998), pp. 1188–1191.

[49] Andrea Wilson. "Bounded Memory and Biases in Information Processing". In: *Econometrica* 82.6 (2014), pp. 2257–2294.

# A Optimality of the processing strategy

Before I present the omitted proofs in the main text, I now discuss the optimality of the equilibrium processing strategy. In the main text, the DM decides whether to process or ignore the signal given the preliminary information about its realization. Therefore, the equilibrium processing strategy $\sigma_t(g, \mathscr{M}_t)$ is optimal for any preliminary information $g$. Moreover, as it maximizes the DM's expected utility for all possible preliminary information in each period, it also maximizes his expected ex-ante utility in every period $t$, which gives the following result.

**Proposition A.1.** *The processing strategy characterized by lemma 1 maximizes the expected utility of the DM's period $t$ self.*

In the maintext, I show that the optimal processing strategy explains some well-documented behavioral phenomena in the presence of limitation in processing ability. The optimality is achieved as the DM evaluates the expected utility of processing and ignoring a signal given his knowledge of the signal realization. In contrast, proposition A.1 suggests that the results also hold if we consider strategies that maximize the DM's expected utility at the beginning of each period. This result is also useful for the proof for the equivalence result, as shown in the next section.

# B Equivalence Results

In this section, I present two equivalence results which allow me to simplify the model with $T > 2$ and $\bar{T} = 1$ to a model with $T = 2$ and $\bar{T} = 1$. The idea is to show that the equilibrium processing strategy at any period $t < T$ in a model with $T > 2$ is equivalent to the equilibrium processing strategy at period 1 in the simplified model with $T = 2$. I first present the result in the unaware case.

## B.1 Unaware Case

**Proposition B.1.** *There exists some p.d.f.* $(\tilde{f}_{2L}, \tilde{f}_{2R})$ *such that the following two equilibrium strategies are equivalent:*

1. *the equilibrium processing strategy of the DM with belief* $(p_L^t, p_R^t)$ *at period* $t < T$*, under a setting with* $T > 2$*,* $\bar{T} = 1$ *and information structure* $\{(f_{t'L}, f_{t'R})\}_{t'=t+1}^T$*;*

2. *the equilibrium processing strategy of the DM with prior belief* $(p_L, p_R) = (p_L^t, p_R^t)$ *at period* $1$*, under a setting with* $T = 2$*,* $\bar{T} = 1$ *and information structure* $(\tilde{f}_{2L}, \tilde{f}_{2R})$*.*

*Proof.* I prove the proposition under the assumption the preliminary information perfectly reveals the signal realization. The proof is similar in the general setting where $g$ could be degenerate[34]. First consider the first point where $T > 2$ and without loss of generality assume $u_L\, p_L^t \geq u_R\, p_R^t$. By lemma 1, the decision maker processes signal $s_t \geq \frac{u_R\, p_R^t}{u_L\, p_L^t}$ if and only if

$$\begin{aligned} s_t &\geq \frac{u_R\, p_R^t}{u_L\, p_L^t} \frac{\triangle_{s_t,\mathscr{M}_t}^t(r \mid R)}{\triangle_{s_t,\mathscr{M}_t}^t(r \mid L)} \\ &= \frac{u_R\, p_R^t}{u_L\, p_L^t} \frac{\sum_{k=t+1}^T \left[\mathbb{1}_{k>t+1} \prod_{h=t+1}^{k-1} \int_0^\infty (\mathbb{1}_{\sigma_h(s')=I}) f_{hR}(s') ds' + \mathbb{1}_{k=t+1}\right] \int_0^{\frac{u_R\, p_R^t}{u_L\, p_L^t}} \mathbb{1}_{\sigma_k(s)=P} f_{kR}(s)\, ds}{\sum_{k=t+1}^T \left[\mathbb{1}_{k>t+1} \prod_{h=t+1}^{k-1} \int_0^\infty (\mathbb{1}_{\sigma_h(s')=I}) f_{hL}(s') ds' + \mathbb{1}_{k=t+1}\right] \int_0^{\frac{u_R\, p_R^t}{u_L\, p_L^t}} \mathbb{1}_{\sigma_k(s)=P} f_{kL}(s)\, ds}. \end{aligned} \tag{B.1}$$

On the other hand, he processes signal $s_t < \frac{u_R\, p_R^t}{u_L\, p_L^t}$ if and only if

$$\begin{aligned} s_t^{-1} &> \frac{u_L\, p_L^t}{u_R\, p_R^t} \frac{\triangle_{s_t,\mathscr{M}_t}^t(l \mid L)}{\triangle_{s_t,\mathscr{M}_t}^t(l \mid R)} \\ &= \frac{u_L\, p_L^t}{u_R\, p_R^t} \frac{1 - \sum_{k=t+1}^T \left[\mathbb{1}_{k>t+1} \prod_{h=t+1}^{k-1} \int_0^\infty (\mathbb{1}_{\sigma_h^*(s')=I}) f_{hL}(s') ds' + \mathbb{1}_{k=t+1}\right] \int_0^{\frac{u_R\, p_R^t}{u_L\, p_L^t}} \mathbb{1}_{\sigma_k^*(s)=P} f_{kL}(s)\, ds}{1 - \sum_{k=t+1}^T \left[\mathbb{1}_{k>t+1} \prod_{h=t+1}^{k-1} \int_0^\infty (\mathbb{1}_{\sigma_h^*(s')=I}) f_{hR}(s') ds' + \mathbb{1}_{k=t+1}\right] \int_0^{\frac{u_R\, p_R^t}{u_L\, p_L^t}} \mathbb{1}_{\sigma_k^*(s)=P} f_{kR}(s)\, ds}. \end{aligned} \tag{B.2}$$

[34] The proof is identical if I replace $\mathbb{1}_{\sigma_h(s)=P}$ by $\int \mathbb{1}_{\sigma_h(s)=P}\, g(s)\, dg$ and $\mathbb{1}_{\sigma_h(s)=I}$ by $\int \mathbb{1}_{\sigma_h(s)=I}\, g(s)\, dg$.

Now consider point 2 where $T = 2$ with information structure $(\tilde{f}_{2L}, \tilde{f}_{2R})$ and prior belief $p_L = p_L^t$. By lemma 1, the decision maker processes signal $s_1 \geq \frac{u_R\, p_R}{u_L\, p_L}$ if

$$s_t \geq \frac{u_R\, p_R}{u_L\, p_L} \frac{\tilde{F}_{2R}\left(\dfrac{u_R\, p_R}{u_L\, p_L}\right)}{\tilde{F}_{2L}\left(\dfrac{u_R\, p_R}{u_L\, p_L}\right)}, \tag{B.3}$$

and he processes signal $s_1 < \frac{u_R\, p_R}{u_L\, p_L}$ if and only if

$$s_t^{-1} > \frac{u_L\, p_L}{u_R\, p_R} \frac{1 - \tilde{F}_{2L}\left(\dfrac{u_R\, p_R}{u_L\, p_L}\right)}{1 - \tilde{F}_{2R}\left(\dfrac{u_R\, p_R}{u_L\, p_L}\right)}. \tag{B.4}$$

By comparing equation (B.1) to equation (B.3), and equation (B.2) to equation (B.4), the equilibrium processing strategy in the two cases are equivalent if and only if

$$\begin{aligned}
\tilde{F}_{2L}\left(\frac{u_R\, p_R}{u_L\, p_L}\right) &= \sum_{k=t+1}^{T}\left[\mathbb{1}_{k>t+1}\prod_{h=t+1}^{k-1}\int_0^{\infty}(\mathbb{1}_{\sigma_h^*(s')=I}) f_{hL}\, ds' + \mathbb{1}_{k=t+1}\right]\int_0^{\frac{u_R\, p_R^t}{u_L\, p_L^t}} \mathbb{1}_{\sigma_k^*(s)=P} f_{kL}(s)\, ds;\\
\tilde{F}_{2R}\left(\frac{u_R\, p_R}{u_L\, p_L}\right) &= \sum_{k=t+1}^{T}\left[\mathbb{1}_{k>t+1}\prod_{h=t+1}^{k-1}\int_0^{\infty}(\mathbb{1}_{\sigma_h^*(s')=I}) f_{hR}\, ds' + \mathbb{1}_{k=t+1}\right]\int_0^{\frac{u_R\, p_R^t}{u_L\, p_L^t}} \mathbb{1}_{\sigma_k^*(s)=P} f_{kR}(s)\, ds.
\end{aligned} \tag{B.5}$$

To prove that there exists p.d.f. $\tilde{f}_{2L}$ and $\tilde{f}_{2R}$ that generates the c.d.f.s evaluated at $(\frac{u_R\, p_R}{u_L\, p_L})$ with values defined in equation (B.5), it remains to prove that

$$\frac{\tilde{F}_{2R}\left(\dfrac{u_R\, p_R}{u_L\, p_L}\right)}{\tilde{F}_{2L}\left(\dfrac{u_R\, p_R}{u_L\, p_L}\right)} > \frac{u_L\, p_L}{u_R\, p_R}.$$

Note that by definition, equation (B.5) implies that

$$\frac{\tilde{F}_{2R}\left(\frac{u_R\, p_R}{u_L\, p_L}\right)}{\tilde{F}_{2L}\left(\frac{u_R\, p_R}{u_L\, p_L}\right)} = \frac{\Pr^{t+1}_{\mathscr{M}_{t+1}}(r \mid R)}{\Pr^{t+1}_{\mathscr{M}_{t+1}}(r \mid L)}.$$

On the other hand, by proposition A.1, the expected utility of the DM at the beginning of period $t+1$ is weakly greater than that if he chooses to process all $s_t$:

$$\begin{aligned}
&u_L\, p_L(1 - \Pr^{t+1}_{\mathscr{M}_{t+1}}(r \mid L)) + u_R\, p_R \Pr^{t+1}_{\mathscr{M}_{t+1}}(r \mid R) \\
\geq & u_L\, p_L \left(1 - F_{(t+1)L}\left(\frac{u_R p_R}{u_L p_L}\right)\right) + u_R\, p_R F_{(t+1)R}\left(\frac{u_R p_R}{u_L p_L}\right) \\
> & u_L\, p_L,
\end{aligned}$$

where the last inequality is implied by the fact receiving one signal always improves expected utility, in comparison to receiving no signal. Rearranging gives:

$$\begin{aligned}
u_R\, p_R \Pr^{t+1}_{\mathscr{M}_{t+1}}(r \mid R) &> u_L\, p_L \Pr^{t+1}_{\mathscr{M}_{t+1}}(r \mid L); \\
\frac{\Pr^{t+1}_{\mathscr{M}_{t+1}}(r \mid R)}{\Pr^{t+1}_{\mathscr{M}_{t+1}}(r \mid L)} &> \frac{u_L\, p_L}{u_R\, p_R}.
\end{aligned}$$

The results follow. □

## B.2 Aware Case

**Proposition B.2.** *There exists some p.d.f. $(\tilde{f}_{2L}, \tilde{f}_{2R})$ such that the following two sets of equilibrium strategies are equivalent:*

1. *the set of equilibrium processing strategies of the DM with belief $(p^t_L, p^t_R)$ at period $t < T$, under a setting with $T > 2$, $\bar{T} = 1$ and information structure $\{(f_{t'L}, f_{t'R})\}^T_{t'=t+1}$ (assuming that it exists);*

2. *the set of equilibrium processing strategies of the DM with prior belief $(p_L, p_R) = (p^t_L, p^t_R)$ at period $1$, under a setting with $T = 2$, $\bar{T} = 1$*

*and information structure* $(\tilde{f}_{2L}, \tilde{f}_{2R})$.

*Proof.* First consider point 1 where $T > 2$ and without loss of generality assume $u_L\, p_L^t \geq u_R\, p_R^t$. For simplicity and with a bit of abuse in notations, define $\mathscr{R}_{t'}^* = \prod_{t=1}^{t'-1} \mathscr{R}_t(\tilde{\sigma}_t)$, which is the conjectured ratio of ignoring all the previous signals in state $L$ over than in state $R$. Note that at the beginning of period $t' > t$, the expected utility of action $l$ over that of action $r$ equals $\frac{u_L\, p_L^t \mathscr{R}_{t'}^*}{u_R\, p_R^t}$. By lemma 1, the decision maker processes signal $s_t \geq \frac{u_R\, p_R^t}{u_L\, p_L^t}$ if and only if

$$\begin{aligned} s_t &\geq \frac{u_R\, p_R^t}{u_L\, p_L^t} \frac{\triangle_{s_t,\mathscr{M}_t}^t(r \mid R)}{\triangle_{s_t,\mathscr{M}_t}^t(r \mid L)} \\ &= \frac{u_R\, p_R^t}{u_L\, p_L^t} \frac{\sum_{k=t+1}^T \left[\mathbb{1}_{k>t+1} \prod_{h=t+1}^{k-1} \int_0^\infty (\mathbb{1}_{\sigma_h(s')=I}) f_{hR}(s') ds' + \mathbb{1}_{k=t+1}\right] \int_0^{\frac{u_R\, p_R^t}{u_L\, p_L^t \mathscr{R}_k^*}} \mathbb{1}_{\sigma_k(s)=P} f_{kR}(s)\, ds}{\sum_{k=t+1}^T \left[\mathbb{1}_{k>t+1} \prod_{h=t+1}^{k-1} \int_0^\infty (\mathbb{1}_{\sigma_h(s')=I}) f_{hL}(s') ds' + \mathbb{1}_{k=t+1}\right] \int_0^{\frac{u_R\, p_R^t}{u_L\, p_L^t \mathscr{R}_k^*}} \mathbb{1}_{\sigma_k(s)=P} f_{kL}(s)\, ds}. \end{aligned} \tag{B.6}$$

On the other hand, he processes signal $s_t < \frac{u_R\, p_R^t}{u_L\, p_L^t}$ if and only if

$$\begin{aligned} s_t^{-1} &> \frac{u_L\, p_L^t}{u_R\, p_R^t} \frac{\triangle_{s_t,\mathscr{M}_t}^t(l \mid L)}{\triangle_{s_t,\mathscr{M}_t}^t(l \mid R)} \\ &= \frac{u_L\, p_L^t}{u_R\, p_R^t} \frac{1 - \sum_{k=t+1}^T \left[\mathbb{1}_{k>t+1} \prod_{h=t+1}^{k-1} \int_0^\infty (\mathbb{1}_{\sigma_h(s')=I}) f_{hL}(s') ds' + \mathbb{1}_{k=t+1}\right] \int_0^{\frac{u_R\, p_R^t}{u_L\, p_L^t \mathscr{R}_k^*}} \mathbb{1}_{\sigma_k(s)=P} f_{kL}(s)\, ds}{1 - \sum_{k=t+1}^T \left[\mathbb{1}_{k>t+1} \prod_{h=t+1}^{k-1} \int_0^\infty (\mathbb{1}_{\sigma_h(s')=I}) f_{hR}(s') ds' + \mathbb{1}_{k=t+1}\right] \int_0^{\frac{u_R\, p_R^t}{u_L\, p_L^t \mathscr{R}_k^*}} \mathbb{1}_{\sigma_k(s)=P} f_{kR}(s)\, ds}. \end{aligned} \tag{B.7}$$

The equilibrium strategy is given by the equation (B.6) and (B.7). Now define $\tilde{s} = s\mathscr{R}_k^*$, note that

$$
\begin{aligned}
&\frac{\sum_{k=t+1}^{T}\left[\prod_{h=t+1}^{k-1}\int_0^{\infty}(\mathbb{1}_{\sigma_h(s')=I})f_{hR}(s')ds'\right]\int_0^{\frac{u_R p_R^t}{u_L p_L^t \mathscr{R}_k^*}}\mathbb{1}_{\sigma_k(s)=P}f_{kR}(s)ds}{\sum_{k=t+1}^{T}\left[\prod_{h=t+1}^{k-1}\int_0^{\infty}(\mathbb{1}_{\sigma_h(s')=I})f_{hL}(s')ds'\right]\int_0^{\frac{u_R p_R^t}{u_L p_L^t \mathscr{R}_k^*}}\mathbb{1}_{\sigma_k(s)=P}f_{kL}(s)ds}\\
=&\frac{\sum_{k=t+1}^{T}\left[\prod_{h=t+1}^{k-1}\int_0^{\infty}(\mathbb{1}_{\sigma_h(s')=I})f_{hR}(s')ds'\right]\int_0^{\frac{u_R p_R^t}{u_L p_L^t \mathscr{R}_k^*}}\mathbb{1}_{\sigma_k(\tilde{s}/\mathscr{R}_k^*)=P}f_{kR}(\tilde{s}/\mathscr{R}_k^*)d\tilde{s}}{\sum_{k=t+1}^{T}\left[\prod_{h=t+1}^{k-1}\int_0^{\infty}(\mathbb{1}_{\sigma_h(s')=I})f_{hL}(s')ds'\right]\int_0^{\frac{u_R p_R^t}{u_L p_L^t \mathscr{R}_k^*}}\mathbb{1}_{\sigma_k(\tilde{s}/\mathscr{R}_k^*)=P}f_{kL}(\tilde{s}/\mathscr{R}_k^*)d\tilde{s}}\\
=&\frac{\int_0^{\frac{u_R p_R^t}{u_L p_L^t \mathscr{R}_k^*}}\sum_{k=t+1}^{T}\left[\prod_{h=t+1}^{k-1}\int_0^{\infty}(\mathbb{1}_{\sigma_h(s')=I})f_{hR}(s')ds'\right]\mathbb{1}_{\sigma_k(\tilde{s}/\mathscr{R}_k^*)=P}f_{kR}(\tilde{s}/\mathscr{R}_k^*)d\tilde{s}}{\int_0^{\frac{u_R p_R^t}{u_L p_L^t \mathscr{R}_k^*}}\sum_{k=t+1}^{T}\left[\prod_{h=t+1}^{k-1}\int_0^{\infty}(\mathbb{1}_{\sigma_h(s')=I})f_{hL}(s')ds'\right]\mathbb{1}_{\sigma_k(\tilde{s}/\mathscr{R}_k^*)=P}f_{kL}(\tilde{s}/\mathscr{R}_k^*)d\tilde{s}}.
\end{aligned}
$$

Now define $(\tilde{f}_{2L}, \tilde{f}_{2R})$ as follows and verify whether it is a probability distribution function and whether the equilibrium processing strategy of the DM in period $t = 1$ where $T = 2$ is equivalent to that characterized in equation (B.6) and (B.7):

$$
\begin{aligned}
\tilde{f}_{2L}(\tilde{s}) &= \sum_{k=t+1}^{T}\left[\prod_{h=t+1}^{k-1}\int_0^{\infty}(\mathbb{1}_{\sigma_h(s')=I})f_{hL}(s')ds'\right]\mathbb{1}_{\sigma_k(\tilde{s}/\mathscr{R}_k^*)=P}f_{kL}(\tilde{s}/\mathscr{R}_k^*);\\
\tilde{f}_{2R}(\tilde{s}) &= \sum_{k=t+1}^{T}\left[\prod_{h=t+1}^{k-1}\int_0^{\infty}(\mathbb{1}_{\sigma_h(s')=I})f_{hR}(s')ds'\right]\mathbb{1}_{\sigma_k(\tilde{s}/\mathscr{R}_k^*)=P}f_{kR}(\tilde{s}/\mathscr{R}_k^*).
\end{aligned}
$$

First, it is clear that $\int \tilde{f}_{2L}(\tilde{s})d\tilde{s} = \int \tilde{f}_{2R}(\tilde{s})d\tilde{s} = 1$ as the DM always processes one signal in equilibrium. On the other hand,

$$\frac{\tilde{f}_{2L}(\tilde{s})}{\tilde{f}_{2R}(\tilde{s})} = \frac{\sum_{k=t+1}^{T}\left[\prod_{h=t+1}^{k-1}\int_0^\infty(\mathbb{1}_{\sigma_h(s')=I})f_{hL}(s')ds'\right]\mathbb{1}_{\sigma_k(\tilde{s}/\mathscr{R}_k^*)=P}f_{kL}(\tilde{s}/\mathscr{R}_k^*)}{\sum_{k=t+1}^{T}\left[\prod_{h=t+1}^{k-1}\int_0^\infty(\mathbb{1}_{\sigma_h(s')=I})f_{hR}(s')ds'\right]\mathbb{1}_{\sigma_k(\tilde{s}/\mathscr{R}_k^*)=P}f_{kR}(\tilde{s}/\mathscr{R}_k^*)}$$
$$= \frac{\sum_{k=t+1}^{T}\mathscr{R}_k^*\left[\prod_{h=t+1}^{k-1}\int_0^\infty(\mathbb{1}_{\sigma_h(s')=I})f_{hR}(s')ds'\right]\mathbb{1}_{\sigma_k(\tilde{s}/\mathscr{R}_k^*)=P}f_{kL}(\tilde{s}/\mathscr{R}_k^*)}{\sum_{k=t+1}^{T}\left[\prod_{h=t+1}^{k-1}\int_0^\infty(\mathbb{1}_{\sigma_h(s')=I})f_{hR}(s')ds'\right]\mathbb{1}_{\sigma_k(\tilde{s}/\mathscr{R}_k^*)=P}f_{kR}(\tilde{s}/\mathscr{R}_k^*)}$$
$$= \frac{\sum_{k=t+1}^{T}\tilde{s}\left[\prod_{h=t+1}^{k-1}\int_0^\infty(\mathbb{1}_{\sigma_h(s')=I})f_{hR}(s')ds'\right]\mathbb{1}_{\sigma_k(\tilde{s}/\mathscr{R}_k^*)=P}f_{kR}(\tilde{s}/\mathscr{R}_k^*)}{\sum_{k=t+1}^{T}\left[\prod_{h=t+1}^{k-1}\int_0^\infty(\mathbb{1}_{\sigma_h(s')=I})f_{hR}(s')ds'\right]\mathbb{1}_{\sigma_k(\tilde{s}/\mathscr{R}_k^*)=P}f_{kR}(\tilde{s}/\mathscr{R}_k^*)}$$
$$= \tilde{s}.$$

Denote $(\tilde{F}_{2L}, \tilde{F}_{2R})$ as the c.d.f. associated with $(\tilde{f}_{2L}, \tilde{f}_{2R})$. In point 2 where $T = 2$, the DM processes signal $s_1 \geq \frac{u_R\,p_R}{u_L\,p_L}$ if and only if

$$s_1 \geq \frac{u_R\,p_R}{u_L\,p_L}\frac{\tilde{F}_{2R}\left(\frac{u_R\,p_R}{u_L\,p_L\mathscr{R}_1^*}\right)}{\tilde{F}_{2L}\left(\frac{u_R\,p_R}{u_L\,p_L\mathscr{R}_1^*}\right)},$$

and processes signal $s_t < \frac{u_R\,p_R}{u_L\,p_L}$ if and only if

$$s_1^{-1} \geq \frac{u_L\,p_L}{u_R\,p_R}\frac{1-\tilde{F}_{2L}\left(\frac{u_R\,p_R}{u_L\,p_L\mathscr{R}_1^*}\right)}{1-\tilde{F}_{2R}\left(\frac{u_R\,p_R}{u_L\,p_L\mathscr{R}_1^*}\right)}.$$

The result follows. □

# C Omitted Results and Proofs

## C.1 Proof of Proposition 1

*Proof.* By lemma 1, the decision maker processes $s_1 \geq \frac{u_R\,p_R}{u_L\,p_L}$ if and only if

$$s_1 \geq \frac{u_R\,p_R}{u_L\,p_L}\,\frac{F_{2R}\left(\frac{u_R\,p_R}{u_L\,p_L}\right)}{F_{2L}\left(\frac{u_R\,p_R}{u_L\,p_L}\right)} = \frac{u_R\,p_R}{u_L\,p_L} \times STR\left(s_2 \in \left(0, \frac{u_R\,p_R}{u_L\,p_L}\right)\right) > 1. \quad \text{(C.1)}$$

The last inequality follows from the fact that the strength of the set of signals $s_2 \in (0, \frac{u_R\,p_R}{u_L\,p_L})$ is higher than the strength of $s_2 = \frac{u_R\,p_R}{u_L\,p_L}$, *i.e.*,

$$\frac{F_{2R}\left(\frac{u_R\,p_R}{u_L\,p_L}\right)}{F_{2L}\left(\frac{u_R\,p_R}{u_L\,p_L}\right)} > \frac{f_{2R}\left(\frac{u_R\,p_R}{u_L\,p_L}\right)}{f_{2L}\left(\frac{u_R\,p_R}{u_L\,p_L}\right)} = \frac{u_L\,p_L}{u_R\,p_R}.$$

The last inequality of equation (C.1) also implies that the DM ignores all weak belief-challenging signals $s_1 \in [\frac{u_R\,p_R}{u_L\,p_L}, 1]$.

On the other hand, the DM processes $s_1 < \frac{u_R\,p_R}{u_L\,p_L}$ if and only if

$$s_1^{-1} > \frac{u_L\,p_L}{u_R\,p_R}\,\frac{1-F_{2L}\left(\frac{u_R\,p_R}{u_L\,p_L}\right)}{1-F_{2R}\left(\frac{u_R\,p_R}{u_L\,p_L}\right)} = \frac{u_L\,p_L}{u_R\,p_R} \times STR\left(s_2 \in \left[\frac{u_R\,p_R}{u_L\,p_L}, \infty\right)\right) > \frac{u_L\,p_L}{u_R\,p_R}. \quad \text{(C.2)}$$

The last inequality is implied by

$$\begin{aligned} F_{2R}\left(\frac{u_R\,p_R}{u_L\,p_L}\right) &> F_{2L}\left(\frac{u_R\,p_R}{u_L\,p_L}\right) \\ 1-F_{2L}\left(\frac{u_R\,p_R}{u_L\,p_L}\right) &> 1-F_{2R}\left(\frac{u_R\,p_R}{u_L\,p_L}\right) \\ STR\left(s_2 \in \left[\frac{u_R\,p_R}{u_L\,p_L}, \infty\right)\right) &> 1. \end{aligned}$$

Combining inequalities (C.1) and (C.2) proves the results. $\square$

## C.2 Proof of Proposition 2

*Proof.* Point 1 of the proposition is directly implied by the proposition 1. For point 2, by proposition 1, $\Phi^- > \Phi^+$ if and only if:

$$\frac{u_L^2\, p_L^2}{u_R^2\, p_R^2} \times \frac{\left(1 - F_{2L}\left(\frac{u_R\, p_R}{u_L\, p_L}\right)\right) F_{2L}\left(\frac{u_R\, p_R}{u_L\, p_L}\right)}{\left(1 - F_{2R}\left(\frac{u_R\, p_R}{u_L\, p_L}\right)\right) F_{2R}\left(\frac{u_R\, p_R}{u_L\, p_L}\right)} > 1.$$

First, note that

$$\lim_{u_L\, p_L/u_R\, p_R \to \infty} \frac{F_{2L}\left(\frac{u_R\, p_R}{u_L\, p_L}\right)}{F_{2R}\left(\frac{u_R\, p_R}{u_L\, p_L}\right)} = \lim_{u_L\, p_L/u_R\, p_R \to \infty} \frac{f_{2L}\left(\frac{u_R\, p_R}{u_L\, p_L}\right)}{f_{2R}\left(\frac{u_R\, p_R}{u_L\, p_L}\right)} = \lim_{u_L\, p_L/u_R\, p_R \to \infty} \frac{u_R\, p_R}{u_L\, p_L}$$

which implies

$$\lim_{u_L\, p_L/u_R\, p_R \to \infty} \frac{u_L\, p_L F_{2L}\left(\frac{u_R\, p_R}{u_L\, p_L}\right)}{u_R\, p_R F_{2R}\left(\frac{u_R\, p_R}{u_L\, p_L}\right)} = 1 > 0;$$

$$\lim_{u_L\, p_L/u_R\, p_R \to \infty} \frac{u_L^2\, p_L^2}{u_R^2\, p_R^2} \times \frac{\left(1 - F_{2L}\left(\frac{u_R\, p_R}{u_L\, p_L}\right)\right) F_{2L}\left(\frac{u_R\, p_R}{u_L\, p_L}\right)}{\left(1 - F_{2R}\left(\frac{u_R\, p_R}{u_L\, p_L}\right)\right) F_{2R}\left(\frac{u_R\, p_R}{u_L\, p_L}\right)} = +\infty > 0.$$

The result follows given the continuity of $\Phi^-$ and $\Phi^+$ in $p_L$. □

## C.3 Proof of Corollary 3

*Proof.* I prove the corollary at the limit where $p_L \to 1$. Follow from the proof of proposition 2,

$$\lim_{u_L p_L / u_R p_R \to \infty} \Phi^+ = 1;$$
$$\lim_{u_L p_L / u_R p_R \to \infty} \Phi^- = +\infty.$$

It is then obvious that as $p_L \to 1$, for all $s > 1$,

$$U_P(s, \mathscr{M}_1) > U_I(s, \mathscr{M}_1)$$
$$U_P(s^{-1}, \mathscr{M}_1) < U_I(s^{-1}, \mathscr{M}_1).$$

The result follows from corollary 2. □

## C.4 Proof of Corollary 4

*Proof.* I prove the corollary at the limit where $p_L \to 1$. Follow from the proof of proposition 2,

$$\lim_{u_L p_L / u_R p_R \to \infty} \Phi^+ = 1;$$
$$\lim_{u_L p_L / u_R p_R \to \infty} \Phi^- = +\infty.$$

Now suppose $\Phi^+ \to 1$ and $\Phi^- \to +\infty$. The probability that the DM chooses action $r$ converges to

$$p_L F_{1L}(1) F_{2L}\left(\frac{u_R p_R}{u_L p_L}\right) + p_R F_{1R}(1) F_{2R}\left(\frac{u_R p_R}{u_L p_L}\right) \tag{C.3}$$

On the other hand, if the DM can process both signals, the probability that he chooses action $r$ equals

$$p_L \int_0^\infty F_{2L}\left(\frac{u_R p_R}{u_L p_L s}\right) f_{1L}(s) ds + p_R \int_0^\infty F_{2R}\left(\frac{u_R p_R}{u_L p_L s}\right) f_{1R}(s) ds. \tag{C.4}$$

Equation (C.3) is strictly smaller than equation (C.4) as for $\omega = L$ and $R$,

$$\begin{aligned}\int_0^\infty F_{2\omega}\left(\frac{u_R\, p_R}{u_L\, p_L s}\right) f_{1\omega}(s)ds &> \int_0^1 F_{2\omega}\left(\frac{u_R\, p_R}{u_L\, p_L s}\right) f_{1\omega}(s)ds \\ &> \int_0^1 F_{2\omega}\left(\frac{u_R\, p_R}{u_L\, p_L}\right) f_{1\omega}(s)ds \\ &= F_{2\omega}\left(\frac{u_R\, p_R}{u_L\, p_L}\right) F_{1\omega}(1).\end{aligned}$$

The result follows by the continuity of the probability of choosing action $r$ in $\Phi^+$ and $\Phi^-$. □

## C.5 Reverse Wishful Thinking

To illustrate reverse wishful thinking, I normalize the utility function as follows:

$$\begin{aligned}u(l \mid L) &= -u_L < 0; \\ u(l \mid R) &= u_R > 0; \\ u(r \mid R) = u(r \mid L) &= 0.\end{aligned}$$

State $L$ is the undesirable outcome which yields weakly negative utility while state $R$ yields weakly positive utility. By analogy to the definition of wishful thinking, the processing strategy of the DM exhibits reverse wishful thinking when $\Phi^+ < \Phi^+$. That is, he processes a larger set of information that supports the undesirable state, compare to that supports the desirable state.

**Corollary 11 (Reverse Wishful thinking).** *When state $L$ is very undesirable, the equilibrium processing strategy of the DM exhibits reverse wishful thinking, i.e., $\Phi^- > \Phi^+$ when $|-u_L/u_R|$ is big enough.*

## C.6 Proof of Proposition 3

*Proof.* By proposition 1,

$$\begin{aligned}\Phi_A^+ &= \frac{u_R\,p_R}{u_L\,p_L} \times \frac{F_{2R}\left(\frac{u_R\,p_R}{u_L\,p_L}\right)}{F_{2L}\left(\frac{u_R\,p_R}{u_L\,p_L}\right)} \\ &= \Phi_B^+\end{aligned}$$

On the other hand,

$$\begin{aligned}\Phi_A^- &= \frac{u_L\,p_L}{u_R\,p_R} \times \frac{1-\lambda_A+\lambda_A\left(1-F_{2L}\left(\frac{u_R\,p_R}{u_L\,p_L}\right)\right)}{1-\lambda_A+\lambda_A\left(1-F_{2R}\left(\frac{u_R\,p_R}{u_L\,p_L}\right)\right)} \\ &> \frac{u_L\,p_L}{u_R\,p_R} \times \frac{1-\lambda_B+\lambda_B\left(1-F_{2L}\left(\frac{u_R\,p_R}{u_L\,p_L}\right)\right)}{1-\lambda_B+\lambda_B\left(1-F_{2R}\left(\frac{u_R\,p_R}{u_L\,p_L}\right)\right)} \\ &= \Phi_B^-\end{aligned} \tag{C.5}$$

where the second inequality of equation (C.5) follows from the fact that $1 - F_{2L}\left(\frac{u_R p_R}{u_L p_L}\right) > 1 - F_{2R}\left(\frac{u_R p_R}{u_L p_L}\right)$. ☐

## C.7 Proof of Proposition 4

*Proof.* First, by proposition 1,

$$
\begin{aligned}
\Phi_A^- &= \frac{u_L p_L}{u_R p_R} \frac{1-\lambda-\delta+\lambda\left(1-F_{2L}^B\left(\frac{u_R p_R}{u_L p_L}\right)\right)+\delta\left(1-F_{2L}^A\left(\frac{u_R p_R}{u_L p_L}\right)\right)}{1-\lambda-\delta+\lambda\left(1-F_{2R}^B\left(\frac{u_R p_R}{u_L p_L}\right)\right)+\delta\left(1-F_{2R}^A\left(\frac{u_R p_R}{u_L p_L}\right)\right)} \\
&> \frac{u_L p_L}{u_R p_R} \frac{1-\lambda+\lambda\left(1-F_{2L}^B\left(\frac{u_R p_R}{u_L p_L}\right)\right)}{1-\lambda+\lambda\left(1-F_{2R}^B\left(\frac{u_R p_R}{u_L p_L}\right)\right)} \\
&= \Phi_B^-
\end{aligned}
\tag{C.6}
$$

where the second inequality of equation (C.6) follows from the fact that $1-F_{2L}^A\left(\frac{u_R p_R}{u_L p_L}\right) > 1-F_{2R}^A\left(\frac{u_R p_R}{u_L p_L}\right)$.

Now I compare $\Phi_A^+$ and $\Phi_B^+$. By proposition 1,

$$
\Phi_A^+ = \frac{u_R p_R}{u_L p_L} \frac{\lambda F_{2R}^B\left(\frac{u_R p_R}{u_L p_L}\right)+\delta F_{2R}^A\left(\frac{u_R p_R}{u_L p_L}\right)}{\lambda F_{2L}^B\left(\frac{u_R p_R}{u_L p_L}\right)+\delta F_{2L}^A\left(\frac{u_R p_R}{u_L p_L}\right)};
$$

$$
\Phi_B^+ = \frac{u_R p_R}{u_L p_L} \frac{\lambda F_{2R}^B\left(\frac{u_R p_R}{u_L p_L}\right)}{\lambda F_{2L}^B\left(\frac{u_R p_R}{u_L p_L}\right)}.
$$

Therefore, $\Phi_A^+ \leq \Phi_B^+$ if and only if

$$
\frac{F_{2R}^A\left(\frac{u_R p_R}{u_L p_L}\right)}{F_{2L}^A\left(\frac{u_R p_R}{u_L p_L}\right)} \leq \frac{F_{2R}^B\left(\frac{u_R p_R}{u_L p_L}\right)}{F_{2L}^B\left(\frac{u_R p_R}{u_L p_L}\right)}.
$$

□

## C.8 Proof of Proposition 5

*Proof.* By proposition B.2, it is sufficient to prove that there exists a perfect Bayesian Nash equilibrium for $T = 2$. First note that the equilibrium strategy is a threshold strategy, with thresholds $(\Phi^+, \Phi^-)$. Denote $\mathscr{R}_1$ as

$$\mathscr{R}_1 = \frac{F_{1L}(\Phi^+) - F_{1L}(1/\Phi^-)}{F_{1R}(\Phi^+) - F_{1R}(1/\Phi^-)}.$$

As $\Phi^+ \geq \frac{u_R p_R}{u_L p_L}$ and $\Phi^- > \frac{u_L p_L}{u_R p_R}$, $\mathscr{R}_1 \in \left(\frac{F_{1L}(1)}{F_{1R}(1)}, \frac{1-F_{1L}(1)}{1-F_{1R}(1)}\right)$. Given the optimality of the processing strategy, the perfect Bayesian Nash equilibrium is a fixed point of the following equation:

$$\mathscr{R}_1 = \frac{F_{1L}\left(\frac{u_R p_R}{u_L p_L}\frac{F_{2R}\left(\frac{u_R p_R}{u_L p_L \mathscr{R}_1}\right)}{F_{2L}\left(\frac{u_R p_R}{u_L p_L \mathscr{R}_1}\right)}\right) - F_{1L}\left(\frac{u_R p_R}{u_L p_L}\frac{1-F_{2R}\left(\frac{u_R p_R}{u_L p_L \mathscr{R}_1}\right)}{1-F_{2L}\left(\frac{u_R p_R}{u_L p_L \mathscr{R}_1}\right)}\right)}{F_{1R}\left(\frac{u_R p_R}{u_L p_L}\frac{F_{2R}\left(\frac{u_R p_R}{u_L p_L \mathscr{R}_1}\right)}{F_{2L}\left(\frac{u_R p_R}{u_L p_L \mathscr{R}_1}\right)}\right) - F_{1R}\left(\frac{u_R p_R}{u_L p_L}\frac{1-F_{2R}\left(\frac{u_R p_R}{u_L p_L \mathscr{R}_1}\right)}{1-F_{2L}\left(\frac{u_R p_R}{u_L p_L \mathscr{R}_1}\right)}\right)}$$

Denote the equation as $\mathscr{R}_1 = \psi(\mathscr{R}_1)$. The mapping $\psi$ is continuous mapping from $\left[\frac{F_{1L}(1)}{F_{1R}(1)}, \frac{1-F_{1L}(1)}{1-F_{1R}(1)}\right]$ to itself, which is a convex and compact set. By the Brouwer's fixed point theorem, there exists a fixed point of the equation which belongs to the set $\left[\frac{F_{1L}(1)}{F_{1R}(1)}, \frac{1-F_{1L}(1)}{1-F_{1R}(1)}\right]$. As the two end points $\frac{F_{1L}(1)}{F_{1R}(1)}$ and $\frac{1-F_{1L}(1)}{1-F_{1R}(1)}$ are not a fix point of the equation, there exists a fixed point $\mathscr{R}_1^*$ such that

$$\mathscr{R}_1^* \in \left(\frac{F_{1L}(1)}{F_{1R}(1)}, \frac{1-F_{1L}(1)}{1-F_{1R}(1)}\right)$$
$$\mathscr{R}_1^* = \psi(\mathscr{R}_1^*).$$

□

## C.9 Proof of Corollary 7

*Proof.* Without loss of generality, I analyze only the processing strategy of Alice. First by corollary 1, Alice processes signal $q$ because

$$q > \frac{1-p}{p} \times q,$$

while she processes signal $q^{-1}$ if and only if

$$q > \frac{p}{1-p} \times \frac{1-\lambda+\lambda q(1+q)^{-1}}{1-\lambda+\lambda(1+q)^{-1}} = \frac{p}{1-p} \times \frac{1-\lambda+q}{1+q(1-\lambda)}.$$

Define $\mathscr{F}(p,q,\lambda) = q - \frac{p}{1-p} \times \frac{1-\lambda+q}{1+q(1-\lambda)}$. Its first derivatives w.r.t. $p$ and $\lambda$ are

$$\begin{aligned}\frac{\partial \mathscr{F}}{\partial p} &= -\frac{1}{(1-p)^2}\frac{1-\lambda+q}{1+q(1-\lambda)} < 0;\\ \frac{\partial \mathscr{F}}{\partial \lambda} &= -\frac{p}{1-p}\frac{q^2-1}{(1+q(1-\lambda))^2} < 0.\end{aligned}$$

Moreover,

$$\begin{aligned}\mathscr{F}(0,q,\lambda) = q > 0 > \lim_{(p/1-p)\to q} \mathscr{F}(p,q,\lambda) = q\frac{\lambda(1-q)}{1+q(1-\lambda)};\\ \mathscr{F}(p,q,0) = q - \frac{p}{1-p} > 0 > \mathscr{F}(p,q,1) = q(1-\frac{p}{1-p}).\end{aligned}$$

Therefore, there exists $p^-$ and $\lambda^-$ such that $\mathscr{F} > 0$ if and only if $p < p^-$ and if and only if $\lambda < \lambda^-$. It remains to prove point 2 of the corollary. First, $\mathscr{F}$ is strictly convex in $q$

$$\frac{\partial^2 \mathscr{F}}{\partial q^2} = \frac{p}{1-p}\frac{2\lambda(1-\lambda)(2-\lambda)}{(1+q(1-\lambda))^3} > 0,$$

and the two roots of $\mathscr{F}(q) = 0$ are

$$\frac{-(1-\frac{p}{1-p}) \pm \sqrt{(1-\frac{p}{1-p})^2 + 4(1-\lambda)^2\frac{p}{1-p}}}{2(1-\lambda)}. \tag{C.7}$$

One of the two roots is negative, which contradicts the fact that $q > 1$. Hence, there exists a $q^-$, which is the positive root defined in equation (C.7), such that $\mathscr{F} > 0$ if and only if $q > q^-$. □

## C.10 Proof of Proposition 8

*Proof.* When $\lambda < \lambda^-$, Alice and Bob processes both belief-confirming and belief-challenging information. They take the same action if and only if at least one of $s_1$ and $s_2$ is informative.

$$P_{concensus} = 1 - (1-\lambda)^2 = \lambda(2-\lambda). \tag{C.8}$$

On the other hand, when $\lambda \geq \lambda^-$, they ignore belief-challenging signals in period 1. Therefore, Alice and Bob take the same action if and only if the signals in both periods support the same state, or if $s_1$ is pure noise while $s_2$ is not.

$$P_{concensus} = \lambda^2(q^2(1+q)^{-2} + (1+q)^{-2}) + \lambda(1-\lambda) \tag{C.9}$$

Similarly, the probability that Alice/Bob takes her/his a priori optimal action follows:

$$P^a_{default} = \begin{cases} 1 - \lambda(1+q)^{-1} - \lambda(1-\lambda)(1+q)^{-1} & \text{if } \lambda < \lambda^- \\ 1 - \lambda^2(1+q)^{-2} - \lambda(1-\lambda)(1+q)^{-1} & \text{if } \lambda \geq \lambda^- \end{cases}$$
$$P^b_{default} = \begin{cases} 1 - \lambda q(1+q)^{-1} - \lambda(1-\lambda)q(1+q)^{-1} & \text{if } \lambda < \lambda^- \\ 1 - \lambda^2 q^2(1+q)^{-2} - \lambda(1-\lambda)q(1+q)^{-1} & \text{if } \lambda \geq \lambda^- \end{cases} \tag{C.10}$$

First, I prove the first part of the proposition. First, notice that equation (C.8) is increasing and convex in $\lambda$. On the other hand, the first and second derivative of equation (C.9) w.r.t. $\lambda$ are given by:

$$\frac{\partial P_{concensus}}{\partial \lambda} = 1 - 2\lambda(1 - (q^2(1+q)^{-2} + (1+q)^2)) > 0$$
$$\frac{\partial P^2_{concensus}}{\partial^2 \lambda} = -2(1 - (q^2(1+q)^{-2} + (1+q)^2)) < 0$$

respectively. The two inequalities are implied by the fact that $q > 1$ and $\lambda \in [0, 1]$.[35] Secondly, equation (C.9) is clearly smaller than equation (C.8) for any given $\lambda$, which implies that there is a downward jump at $\lambda = \tilde{\lambda}$.

Now I move on to the proof for $P^j_{default}$, $j = a, b$. First, part 1 is implied by the fact that both functions

$$P^a_{default} = 1 - \lambda(1+q)^{-1} - \lambda(1-\lambda)(1+q)^{-1}$$
$$P^b_{default} = 1 - \lambda q(1+q)^{-1} - \lambda(1-\lambda)q(1+q)^{-1}$$

decrease in $\lambda$ as $\lambda(2-\lambda)$ increases in $\lambda$ for $\lambda \in [0, 1]$. On the other hand, from equation (C.10), it is obvious that there is a upward jump at $\lambda = \tilde{\lambda}$. □

[35] First, $[q^2(1+q)^{-2} + (1+q)^{-2}] \in (1/2, 1)$ as it is increasing in $q$. Therefore, $1-(q^2(1+q)^{-2} + (1+q)^{-2}) \in (0, 1/2)$. Combined with the fact that $\lambda \in [0, 1]$, it implies the first inequality.