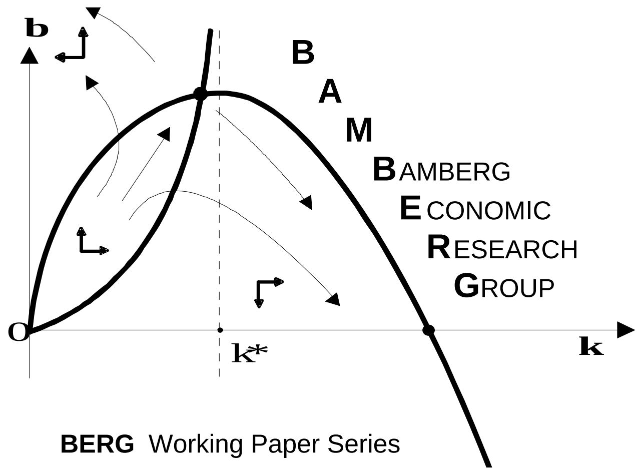
Rational Allocation of Attention in Decision-Making

Stefanie Yvonne Schmitt

Working Paper No. 114

July 2016



Bamberg Economic Research Group
Bamberg University
Feldkirchenstraße 21
D-96052 Bamberg
Telefax: (0951) 863 5547
Telephone: (0951) 863 2687
felix.stuebben@uni-bamberg.de
http://www.uni-bamberg.de/vwl/forschung/berg/

ISBN 978-3-943153-33-0

Redaktion:
Dr. Felix Stübben*

* felix.stuebben@uni-bamberg.de

Rational Allocation of Attention in Decision-Making

Stefanie Yvonne Schmitt*

This version: July 29, 2016

Abstract

This paper proposes a model of attention allocation in decision-making. Attention has various definitions across the literature. Here, I understand attention as selecting information for costly processing. The paper investigates how an agent rationally allocates attention. The resulting attention allocation is context-dependent and influences choice quality. Next to inattention, two strategies of allocating attention prevail. These strategies share similarities with *bottom-up* and *top-down* attention—concepts reported in the psychological literature. Exploring firms' strategic considerations reveals an incentive for firms to produce high quality and highlight quality, if consumers expect low quality, and to exploit consumers by producing low quality and shrouding quality, if agents expect high quality.

KEYWORDS: rational attention, information-processing, decision-making, shrouding.

JEL CODES: D010, D03, D81, D83, L15.

*Department of Economics and Social Sciences, University of Bamberg, Feldkirchenstr. 21, 96045 Bamberg, Germany, stefanie.schmitt@uni-bamberg.de. I am grateful to Florian Herold, Julia Graf, Lisa Planer-Friedrich, and participants at the 5th World Congress of the Game Theory Society (GAMES 2016), the 12th European Meeting on Game Theory (SING12), the GSDS Symposium 2016, economic theory workshops at the University of Bamberg and the 19th BGPE Research Workshop for valuable comments and suggestions. The project has received funding from the European Union's Seventh Framework Programme for research, technological development and demonstration under grant agreement no PCIG11-GA-2012 322253 (Limited Attention).

1 Introduction

This paper proposes a model of attention allocation in decision-making. Across the literature attention has different definitions. Here I understand attention as selecting information for costly processing. The paper investigates how an agent rationally allocates attention to information relevant for a decision. In particular, the paper examines the impact of processing costs on the resulting attention allocation and on the quality of decision-making. Furthermore, taking the resulting attention allocation as a premise of how consumers allocate attention, the paper analyzes firms' optimal response. Specifically, the paper analyzes whether firms use their knowledge about when consumers are inattentive to exploit them.

The ultimate objective of the agent is to choose between two options. This decision is either more important or less important. The importance of a decision is given by the utility difference between the options. In a more important decision more utility is at stake. Independent of the importance of the decision, one of the two options is always better. So two variables influence the utilities of the two options: One variable describes the importance and one—the decisive—variable describes which option is better. The agent knows neither the importance nor which option is better. However, the agent can inform herself by paying attention, i.e., selecting the variables for costly processing. If the agent has not selected a variable for processing, she has to choose between the options given the expected value of that variable. By allocating attention to a variable, the agent eliminates all uncertainty about the realization of that variable.

The following decision exemplifies one kind of situation analyzed in this model: A consumer has to decide between buying (option 1) and not buying (option 2) an umbrella. The umbrella can have high or low quality (in relation to its price). If the quality is high, the consumer prefers to buy the umbrella. If the quality is low, the consumer prefers not to buy the umbrella. The utility difference between buying and not buying is not constant; in some situations the decision is more important. For instance, the decision is more important when it is raining. If the umbrella is of high quality, the consumer always has a benefit from owning the umbrella, but if it rains the consumer has additional benefits from not getting wet immediately after leaving the shop. In contrast, if the umbrella's quality is low (e.g., it is not very robust), the consumer incurs a loss from paying for an useless umbrella. But if it rains the consumer also has an additional loss by getting wet, although she had hoped to be protected from the rain. In this example the importance of the decision is exogenous to the agent. However, it can also be agent-specific, such as a preference for quality. For instance, when choosing a hotel, some are more sensitive about quality (especially housekeeping) than others.

The agent has to decide which variable to pay attention to, before she chooses an option. Thus the model consists of a two-stage decision problem, illustrated in Figure 1. In the first decision the agent chooses the variables she wants to process and the processing order. A selection and corresponding order constitute an *attention strategy*. The set of attention strategies includes all possible ways of selecting the two variables θ and s for processing: Select none, only one, both, or select conditional, e.g., first check how important the decision is, and then—only if the decision is important—check which option is better. After the agent chooses an attention strategy, she processes the selected variable as specified in the chosen attention strategy. This is costly. In the second decision the agent chooses between the available options given the previously processed information. Agents are always rational in the sense that they maximize their expected utility when choosing an attention strategy and when choosing an option.

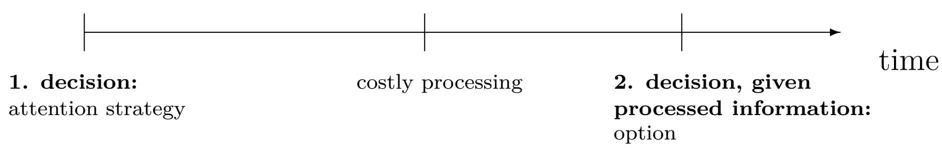


Figure 1: Timeline of the decision-making process.

I do not assume that attention is costly. Attention is only the selection of information for processing. Costly is the actual processing of information, i.e., the consequences of attention. If information-processing were for free, the decision about how to allocate attention would be trivial. A rational agent would gather all pieces of information, calculate the utilities of both options precisely, and always choose the option with the highest utility. However, at the very least the processing of information entails opportunity costs. Experimental and empirical evidence indicates that the costs for processing information are not zero and that these costs have an impact on decisions. Chetty, Looney, and Kroft (2009), for instance, find that—although consumers are mostly aware of sales taxes if asked—if it becomes easier to process the (tax inclusive) price of a good, consumption behavior changes; less is bought.[1] I assume that the processing costs are additive separable to the utility derived from the choice between options 1 and 2.

Results show that three attention strategies prevail: Complete inattention, a direct and exclusive selection of the decisive variable, i.e., the variable that determines which option

[1]Following the psychological literature which assumes a capacity limit on processing information (see e.g., Nobre and Kastner, 2014a, for a brief overview), the costs for processing some variable represent the inability to process other information.

is better, and a conditional attention strategy. With this conditional attention strategy, the agent processes the decisive variable only if the situation is important. Generally, the model shows that attention is context-dependent, i.e., the selection of an attention strategy depends on the processing costs. If the costs for processing the decisive variable are low, the agent processes the variable directly. The agent then chooses between the options knowing which option is better. If the costs for processing the decisive variable are high, the agent stays inattentive and decides given expectations. If the costs for processing the decisive variable are intermediate, such that the benefits of knowing which option is better exceed the costs of processing in the high-stake, but not in the low-stake situation, the agent selects the conditional attention strategy and processes the decisive variable only in the high-stake situation. The always fully informed agent only shows up in the limiting case when processing is costless.

The resulting attention allocation shares similarities with *top-down* and *bottom-up attention*—concepts reported in the psychological literature. Top-down attention describes an endogenous selection of information that requires more cognitive effort. In contrast, bottom-up attention describes an exogenous selection where the salience of the information draws attention automatically.[2] I argue that the conditional selection shares similarities with top-down and the direct selection with bottom-up attention. That the rational considerations of the agent result in the use of attention strategies that share similarities with top-down and bottom-up attention offers an explanation for the existence of top-down and bottom-up attention. Furthermore, processing costs alone—or whether a variable is salient—do not determine whether a variable is processed. In the model only the decisive variable is directly processed. Importance is never directly processed, even if it is very easy to process (i.e., when it is salient). The agent only processes importance with the higher aim to condition the processing of the decisive variable on importance. Thus processing in the model accounts for type and relevance of a variable.

In addition, my model predicts changes in decision-making, such as choice reversals, as a consequence of changing processing costs. Processing costs control which attention strategy is optimal and thus whether decisions are made with information or with expectations. As processing costs change, another attention strategy may be optimal and thus the decision between the options made with information instead of expectations (if costs drop) or vice versa (if costs increase). This is in line with evidence that shows that salience affects behavior (e.g., Chetty, Looney, and Kroft, 2009; Finkelstein, 2009).

Although agents allocate attention rationally, analyzing the optimal response of profit-maximizing firms shows that firms have means to exploit the consumers. Generally, a firm has

[2]See Nobre and Kastner (2014b) for an overview, and especially Nobre and Mesulam (2014).

an incentive to behave contrary to expectations. If consumers expect low quality goods, the firm maximizes profits by producing high quality and highlighting quality. Consumers profit from this: They process the quality and make the optimal decision. Yet, if consumers expect high quality, the firm maximizes profits by producing low quality and shrouding quality by making it costly to process, so that consumers are inattentive to quality. Consumers then decide given the expected quality instead of the true quality. They buy expecting a high quality good but receive a low quality good; their attention allocation is exploited. An equilibrium exists in which a fraction of the firms produce high quality and highlight quality and a fraction produce low quality and shroud quality. The distribution of quality in the market is such that the expected utility from buying equals the utility from not buying.

The rest of the paper is structured as follows: Section 2 integrates the model into the literature on attention and elaborates on differences to existing approaches of modeling attention. Section 3 introduces the model. Section 4 presents the results, discusses implications, and connects the findings to the psychological literature. Section 5 investigates the strategic considerations of firms and section 6 concludes.

2 Related Literature

My model is motivated by empirical and experimental evidence of limited attention: First of all, individuals do not always attend to all available information when they make a decision. For example, consumers are inattentive with regard to add-on costs, such as shipping costs (Hossain and Morgan, 2006; Brown, Hossain, and Morgan, 2010), or with regard to taxes (Chetty, Looney, and Kroft, 2009). Secondly, processing costs are relevant. If processing costs drop, decision-makers are more likely to include the information in their decision. Chetty, Looney, and Kroft (2009), for instance, show that salient taxes impair consumption more than non-salient taxes.

A growing literature attempts to model limited attention. Two strands of that literature are related to my model: General models of attention, particularly models assuming rational attention allocation, and models emphasizing firms' strategic considerations about influencing attention of their consumers. Models of attention are extremely varied; highlighting different aspects and methods of allocating attention. One strand of the literature, for example, focuses on the formation of consideration sets (see, e.g., Manzini and Mariotti, 2014; Eliaz and Spiegler, 2011). This strand assume that the agents are not aware of the existence of all options, but of those they are aware of (consideration set) they know the utility. In contrast, I assume that the agents are aware of the existence of the goods, but are uninformed about the utility.

Assuming rational allocation of attention features prominently in the literature. Gabaix (2014), for instance, analyzes which dimensions of a decision problem an agent takes into account when she decides between actions. My model differs in the setup of costs and in the objective. I focus on the rational attention allocation and set this in relation to psychological attention mechanisms.[3] In addition, I highlight firm's optimal response. In contrast, Gabaix (2014) focuses on an application of his theory in consumption theory and general equilibrium. Another prominent version of modeling rational allocation of attention is rational inattention. The rational inattention literature (e.g., Sims, 2003) assumes that agents maximize utility under a capacity constraint that limits information-processing. Rational inattention utilizes a very specific formulation of information-processing: Reduction of uncertainty modeled as entropy. Thus attention is continuous. In contrast, I capture capacity only indirectly via processing costs and model attention as binary.

In section 5, I introduce firms that influence the processing costs of the consumer, i.e., whether to shroud information. Gabaix and Laibson (2006) and Heidhues, Köeszegi, and Murooka (2016) also analyze shrouding decisions of firms. However, they look at goods that consist of base-good prices and add-on prices. Consequently, shrouding in Gabaix and Laibson (2006) and Heidhues, Köeszegi, and Murooka (2016) means hiding add-on prices and unshrouding by one firm means consumers become aware of all firms' add-on prices. In contrast, I assume that firms shroud quality and cannot unshroud the quality of other firms. Thus in my model an equilibrium can exist in which some but not all firms quality is shrouded. Similarly to Heidhues, Köeszegi, and Murooka (2015), I also find that products that generate negative utility evoke shrouding, because consumers would not buy these goods otherwise. In contrast to my model, both models assume that any distinction in informed and uninformed consumers—next to unshrouding—is exogenously given.

My paper contributes to the literature by proposing a new model of rational attention allocation where agents choose between all possible ways of selecting information for processing and are able to condition attention on the importance of the decision. The paper combines rational considerations with processing costs (e.g., salience). Thereby the model provides a link to observed attention mechanisms, which allows a founded discussion of the implications of the model. The model also supports evidence on the role of salience and evidence of limited attention in decision-making.

[3] Gabaix (2014) also argues that a relationship between psychology and his model exist. Yet, he assumes that in his two stage approach, the first stage, setting attention optimally, relates to Kahneman's sytem 1—a bottom-up approach—and the second stage, i.e., choosing an option, relates to system 2—a top-down approach. In contrast, I assume that in the first stage, when an attention strategy is chosen, the attention strategies differ in that some are top-down and others bottom-up.

3 The Model

This section presents a model of rational allocation of attention in decision-making. The objective of an agent is to choose between two options—option 1 and option 2. The utility of option 2 is normalized to 0, whereas the utility of option 1 is determined by two random variables:

$$u_1 = \theta s$$
$$u_2 = 0,$$

where $\theta \in \{\theta^L, \theta^H\}$ with $0 < \theta^L < \theta^H$ and $s \in \{s^-, s^+\}$ with $s^- < 0 < s^+$. The probability of $\theta \in \{\theta^L, \theta^H\}$ is $p_\theta \in [0, 1]$, where $p_{\theta^L} + p_{\theta^H} = 1$. Henceforth, for notational simplicity, I will refer to p_{θ^H} as p_θ and to p_{θ^L} as $1 - p_\theta$. The probability of $s \in \{s^-, s^+\}$ is $p_s \in [0, 1]$, where $p_{s^-} + p_{s^+} = 1$. Henceforth, for notational simplicity, I will refer to p_{s^+} as p_s and to p_{s^-} as $1 - p_s$. θ and s are independently distributed. Let the expected values of θ and s be $E[\theta] \equiv p_\theta \theta^H + (1 - p_\theta)\theta^L$ and $E[s] \equiv p_s s^+ + (1 - p_s)s^-$. As θ is always positive, s single-handedly determines whether the utility of option 1 is positive or negative. If $s = s^-$, the utility of option 1 is negative. Thus the agent prefers option 2. If $s = s^+$, the utility of option 1 is positive. Thus the agent prefers option 1. The variable θ represents the importance of the decision problem: If $\theta = \theta^H$ the difference between u_1 and u_2 is ceteris paribus greater than if $\theta = \theta^L$. In short: s determines the sign and θ the scale.

I assume that the agent knows the setup of the decision problem (i.e., the probabilities, possible values of the random variables, u_2, etc.). Yet, the agent does not automatically process the realizations of the variables. Only if the agent pays attention to a variable, does she process the realization of that variable. I understand attention as selecting variables for costly processing. Let $c_\theta > 0$ describe the costs of processing θ and $c_s > 0$ the costs of processing s.

Thus the agent has to decide which variable(s) to pay attention to and then processes these variables, before she decides between options 1 and 2.[4] This decision about the allocation of attention builds the corner stone of the model. Different ways of paying attention exist in this decision problem. Each unique way constitutes an *attention strategy*. An *attention strategy* consists of a selection of variables for processing and the corresponding order in which the selected variables will be processed. The situation is illustrated by the decision tree in Figure 2. The tree specifies the possible processing paths. At the root, the agent

[4]The model does not capture repeatedly choosing between the same options, so that the values of θ and s can be inferred from past experience. For example, the model is more suited for explaining phenomena on the durable goods market, than the non-durable goods market.

either starts by processing s, by processing θ, or by processing neither, $(\not\theta, \not s)$. If the agent chooses to start by processing s or θ, the agent can then decide whether to process the other variable. However, as each variable can take on one of two realizations, an attention strategy has to specify whether the other variable is processed for both possible realizations. The rectangular nodes specify nodes where the agent can choose between actions. The circular nodes specify situations where nature chooses one of two realizations. Which realization is given by the probabilities.

Each unique path through the tree accounting for nature by including an action for each contingency constitutes an attention strategy. Thus an attention strategy specifies an action for sequential nodes. Consequently, the chosen attention strategy determines which variables the agent knows at the time she chooses between the two options. The set of attention strategies includes all ways of deterministically selecting the variables for processing. I write an attention strategy as an ordered pair where the order specifies the sequence of processing the variables. In the defined decision problem the set of attention strategies is thus

$$A = \{(s, (\not\theta, \not\theta)), (s, (\not\theta, \theta)), (s, (\theta, \not\theta)), (\theta, s),$$
$$(\not\theta, \not s), (\theta, (\not s, \not s)), (\theta, (\not s, s)), (\theta, (s, \not s))\}$$

where a normally written variable (θ and s) denotes that the agent selects the variable for processing and a canceled out variable ($\not\theta$ and $\not s$) denotes that the agent does not select the variable for processing.

With respect to the tree, actually two strategies exist in which the agent always processes θ and s and thus always has costs for processing θ and for processing s: $(\theta, (s, s))$ and $(s, (\theta, \theta))$. As both strategies describe the same processing and yield equal payoffs, I replace both strategies by (θ, s). Table 1 summarizes the eight attention strategies and the corresponding processing costs.

Attention Strategy	**Attention to**	**Costs**
$(s, (\not\theta, \not\theta))$	s	c_s
$(s, (\not\theta, \theta))$	s and θ if $s = s^-$	$c_s + (1 - p_s)c_\theta$
$(s, (\theta, \not\theta))$	s and θ if $s = s^+$	$c_s + p_s c_\theta$
(θ, s)	θ and s	$c_\theta + c_s$
$(\not\theta, \not s)$	—	—
$(\theta, (\not s, \not s))$	θ	c_θ
$(\theta, (\not s, s))$	θ and s if $\theta = \theta^L$	$c_\theta + (1 - p_\theta)c_s$
$(\theta, (s, \not s))$	θ and s if $\theta = \theta^H$	$c_\theta + p_\theta c_s$

Table 1: Attention strategies and corresponding processing costs.

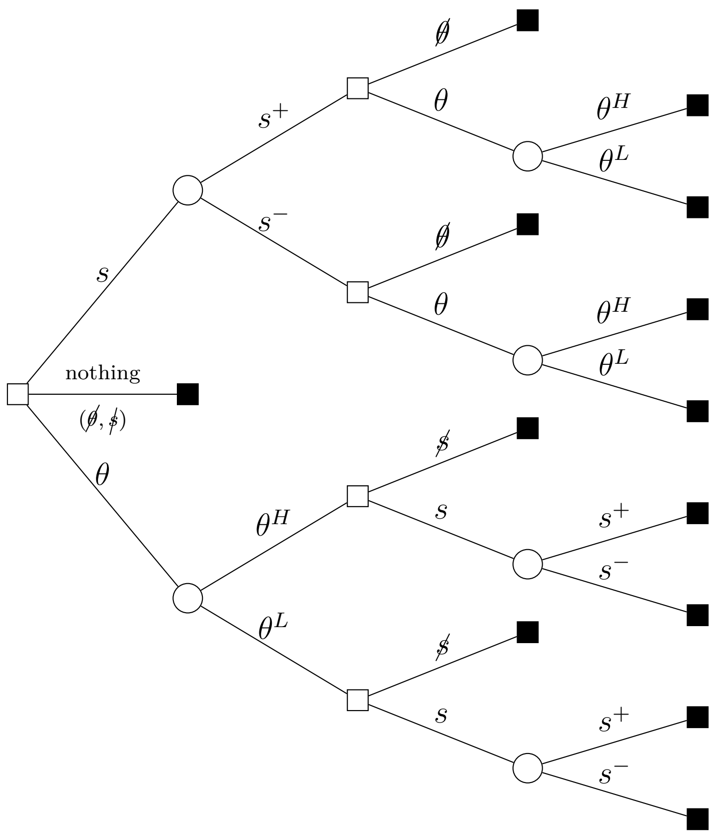


Figure 2: The tree illustrates possible sequences of how the agent can process θ and s. At a rectangular node, the agent decides which action to take. Actions are either to process a variable (θ or s) or not to process a variable ($\not\theta$ or $\not s$). At a circular node the agent processes one of two realizations. The path from the root to an (black) end node symbolizes which information the agent has when she chooses between options 1 and 2. In a representation of the complete decisions of the agent, at the black rectangular nodes the agent would decide between options 1 and 2.

The agent has to choose between the eight attention strategies, before she chooses between the two options. Consequently, the model consists of a two-stage decision problem. I assume that the agent is rational in the sense that she maximizes her expected utility in both choices taking the processing costs into account. At the beginning the agent is aware of the setup, but uncertain about θ and s. In the first decision the agent chooses an attention strategy. Given her selection, she executes the attention strategy by processing the selected variable(s) as specified in the attention strategy. That is, the agent makes her way through the tree according to the attention strategy she chose. In the second decision she chooses an option. In the tree this decision would follow at the black end nodes. As the agent might not attend to both variables, the perception of the utility of option 1 might be fragmentary.

The agent chooses between attention strategies by maximizing expected utility. Thus the expected utilities need to be derived to infer the attention allocation of the agent. Therefore, I elaborate in the following on the various attention strategies and the utility the agent can expect with each strategy.[5] The utility an agent expects with a strategy depends on the option she is going to choose and the processing costs she is going to incur. With attention strategy $(s, (\emptyset, \emptyset))$, the agent only processes s. Because s alone determines which option is better, the agent always knows which option is better and thus chooses the option with the higher utility: If $s = s^+$ she chooses option 1 and if $s = s^-$ she chooses option 2. Nevertheless, she always incurs the costs for processing s. The expected utility is thus

$$E_{(s,(\emptyset,\emptyset))} = E[\theta]p_s s^+ - c_s.$$

Attention strategy $(s, (\emptyset, \theta))$ specifies a conditional allocation of attention. The agent first processes s. If $s = s^+$, she does not process θ. If $s = s^-$, she processes θ. Therefore, the agent always has costs for processing s, but only has costs for processing θ with probability $1 - p_s$. Because the agent always processes s, she always knows which option is better:

$$E_{(s,(\emptyset,\theta))} = E[\theta]p_s s^+ - c_s - (1 - p_s)c_\theta.$$

With attention strategy $(s, (\theta, \emptyset))$, the agent allocates attention conditionally as well. Similarly to strategy $(s, (\emptyset, \theta))$, the agent always processes s. But, she processes θ only if $s = s^+$. Therefore,

$$E_{(s,(\theta,\emptyset))} = E[\theta]p_s s^+ - c_s - p_s c_\theta.$$

If the agent chooses attention strategy (θ, s), she will process all available information and

[5]For a more detailed derivation of the expected utilities of the strategies see Appendix A.

will be able to calculate u_1 precisely. Therefore, she always makes the utility-maximizing choice—independent of the realization of θ. However, she also has maximal perception costs. Therefore, the expected utility of strategy (θ, s) is:

$$E_{(\theta,s)} = E[\theta]p_s s^+ - c_\theta - c_s.$$

If the agent chooses attention strategy $(\not\theta, \not s)$, she processes neither θ nor s. Thus the agent chooses between options 1 and 2 given her expectations about s. If $E[s] < 0$, she chooses option 2. If $E[s] = 0$, she is indifferent and randomizes. And if $E[s] > 0$, she chooses option 1. Compared to all other strategies being inattentive is the cheapest: As the agent processes nothing, she has no processing costs. The expected utility of $(\not\theta, \not s)$ is, therefore,

$$E_{(\not\theta,\not s)} = \begin{cases} 0 & \text{if } E[s] \leq 0 \\ E[\theta]E[s] & \text{if } E[s] > 0. \end{cases}$$

With attention strategy $(\theta, (\not s, \not s))$ the agent allocates attention only to θ. The agent thus never knows which option is better and chooses given her expectations about s. Because she processes θ, she incurs the corresponding costs:

$$E_{(\theta,(\not s,\not s))} = \begin{cases} -c_\theta & \text{if } E[s] \leq 0 \\ E[\theta]E[s] - c_\theta & \text{if } E[s] > 0. \end{cases}$$

Attention strategy $(\theta, (\not s, s))$ also represents a conditional allocation of attention. The agent first processes θ. If $\theta = \theta^H$, she does not process s and has to decide between the options given the expected value of s. If $\theta = \theta^L$, she processes s and thus knows which option yields higher utility. Therefore, the agent always has costs for processing θ, but only has costs for processing s with probability $1 - p_\theta$:

$$E_{(\theta,(\not s,s))} = \begin{cases} (1-p_\theta)p_s\theta^L s^+ - c_\theta - (1-p_\theta)c_s & \text{if } E[s] \leq 0 \\ (1-p_\theta)p_s\theta^L s^+ + p_\theta\theta^H E[s] - c_\theta - (1-p_\theta)c_s & \text{if } E[s] > 0. \end{cases}$$

Attention strategy $(\theta, (s, \not s))$ constitutes a conditional allocation of attention. Similarly to strategy $(\theta, (\not s, s))$, the agent always processes θ. Yet, she processes s only if $\theta = \theta^H$. Therefore,

$$E_{(\theta,(s,\not s))} = \begin{cases} p_\theta p_s\theta^H s^+ - c_\theta - p_\theta c_s & \text{if } E[s] \leq 0 \\ p_\theta p_s\theta^H s^+ + (1-p_\theta)\theta^L E[s] - c_\theta - p_\theta c_s & \text{if } E[s] > 0. \end{cases}$$

4 Results

Each of the eight attention strategies describes a unique way of paying attention. But, not all attention strategies are optimal.

Proposition 1 *Attention strategies* (θ, s)*,* $(\theta, (\not s, \not s))$*,* $(\theta, (\not s, s))$*,* $(s, (\theta, \not\theta))$*, and* $(s, (\not\theta, \theta))$ *are never optimal.*

All proofs are in the appendix.

The intuition behind the proof is that, compared to other strategies, the strategies listed in Proposition 1 entail higher processing costs without increasing the quality of the choice. First, the agent does worse with strategies (θ, s), $(s, (\theta, \not\theta))$, and $(s, (\not\theta, \theta))$ than with strategy $(s, (\not\theta, \not\theta))$.[6] Due to the selection of s in all four strategies, the agent always knows which option is better and makes the utility-maximizing choice with all four strategies. But in contrast to $(s, (\not\theta, \not\theta))$, the agent incurs additional costs for processing θ with strategies (θ, s), $(s, (\theta, \not\theta))$, and $(s, (\not\theta, \theta))$. Second, the agent does worse with strategy $(\theta, (\not s, \not s))$ than with strategy $(\not\theta, \not s)$. Neither strategy includes s for processing. Consequently, the agent chooses between options 1 and 2 given the expected value of s. But with strategy $(\theta, (\not s, \not s))$ the agent has additional costs for processing θ. Even if processing θ is extremely cheap ($c_\theta \longrightarrow 0$), the agent has no incentive to use strategies (θ, s), $(\theta, (\not s, \not s))$, $(s, (\theta, \not\theta))$, or $(s, (\not\theta, \theta))$.[7] Third, although strategy $(\theta, (\not s, s))$ is not strictly inferior to one particular strategy, it is always strictly inferior to at least one other strategy.

Proposition 1 excludes allocation of attention to redundant or useless information. As strategies (θ, s), $(\theta, (\not s, \not s))$, $(\theta, (\not s, s))$, $(s, (\theta, \not\theta))$, and $(s, (\not\theta, \theta))$ are never optimal, a rational agent will never use them. Thus Proposition 1 allows the exclusion of five attention strategies from the analysis. Three strategies prevail: Inattention, $(\not\theta, \not s)$; direct selection of s, $(s, (\not\theta, \not\theta))$; and a conditional allocation, where the agents selects s only if the utility difference between options 1 and 2 is significant, $(\theta, (s, \not s))$. Each of the three strategies is strictly optimal for some cost combinations.

[6] Except when $p_s = 0$ and $p_s = 1$. If $p_s = 0$, $(s, (\theta, \not\theta))$ does as well as $(s, (\not\theta, \not\theta))$, because it represents the same strategy—namely always processing s and never θ. Yet, if $p_s = 0$, $E[s] < 0$ and $(\not\theta, \not s)$ does better than $(s, (\theta, \not\theta))$. So $(s, (\theta, \not\theta))$ is never optimal.

If $p_s = 1$, $(s, (\not\theta, \theta))$ does as well as $(s, (\not\theta, \not\theta))$, because it represents the same strategy—namely always processing s and never θ. Yet, if $p_s = 1$, $E[s] > 0$ and $(\not\theta, \not s)$ does better than $(s, (\not\theta, \theta))$. So $(s, (\not\theta, \theta))$ is never optimal.

[7] Only if $c_\theta = 0$ does the agent actively consider strategies (θ, s), $(\theta, (\not s, \not s))$, $(s, (\not\theta, \theta))$, and $(s, (\theta, \not\theta))$; but even then these strategies do only as well as strategies $(s, (\not\theta, \not\theta))$ and $(\not\theta, \not s)$.

Proposition 2 *There exists a* $\underline{c}_\theta \geq 0$ *such that*

I) for all $c_\theta \geq \underline{c}_\theta$, *there exists a* $\hat{\kappa}$ *such that*

i) if $c_s < \hat{\kappa}$, *strategy* $(s, (\not\theta, \not\theta))$ *is strictly optimal,*

ii) if $c_s > \hat{\kappa}$, *strategy* $(\not\theta, \not s)$ *is strictly optimal,*

II) for all $c_\theta < \underline{c}_\theta$, *there exists a* $\underline{\kappa}$ *and a* $\overline{\kappa}$ *such that*

i) if $c_s < \underline{\kappa}$, *strategy* $(s, (\not\theta, \not\theta))$ *is strictly optimal,*

ii) if $\underline{\kappa} < c_s < \overline{\kappa}$, *strategy* $(\theta, (s, \not s))$ *is strictly optimal,*

iii) if $c_s > \overline{\kappa}$, *strategy* $(\not\theta, \not s)$ *is strictly optimal.*

The proof and the specifications of $\underline{c}_\theta$, $\hat{\kappa}$, $\underline{\kappa}$, and $\overline{\kappa}$ are in the appendix. Figure 3 illustrates Proposition 2 graphically.

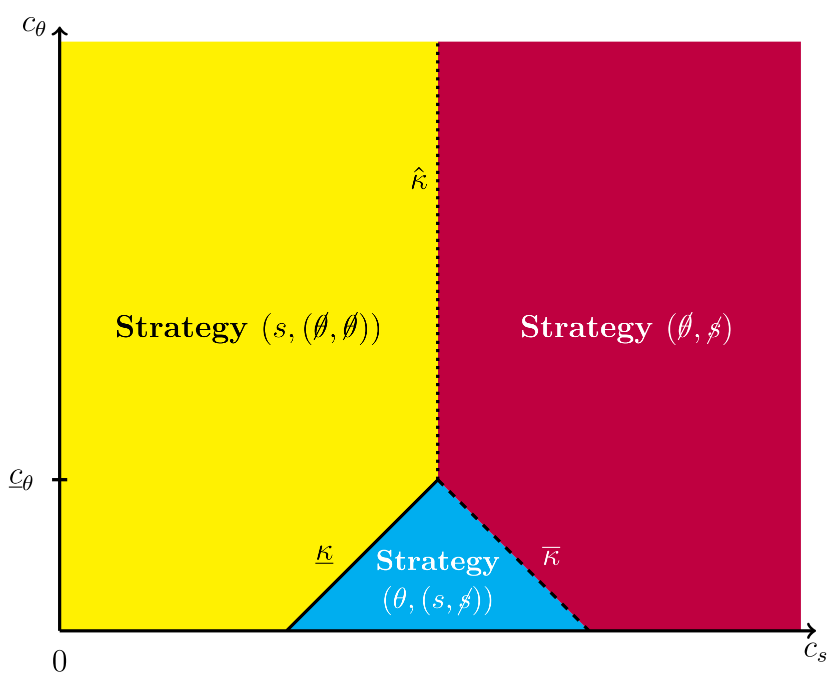


Figure 3: Illustration of Proposition 2. The optimal strategy depends on c_s and c_θ. The black line represents $\underline{\kappa}$. The dashed line represents $\overline{\kappa}$. The dotted line represents $\hat{\kappa}$.

Proposition 2 reveals when each attention strategy is optimal. If the costs for processing θ exceed the threshold $\underline{c}_\theta$, no strategy with which θ is processed can be optimal. The agent chooses either $(s, (\not\theta, \not\theta))$ or $(\not\theta, \not s)$. If the benefits of knowing s exceed the costs of processing s ($c_s < \hat{\kappa}$), processing s directly is optimal, $(s, (\not\theta, \not\theta))$. But if the costs exceed the benefits, it is optimal not to process s, $(\not\theta, \not s)$.

Yet, if $c_\theta < \underline{c}_\theta$, all three strategies prove optimal. The costs for processing s determine which is optimal at any given point. If the costs are low (below threshold $\underline{\kappa}$), strategy $(s, (\not\theta, \not\theta))$—to process s directly—is optimal. If the costs for processing s are high (above threshold $\overline{\kappa}$), strategy $(\not\theta, \not s)$—inattention—is optimal. If the costs are intermediate (costs are between the two thresholds: $\underline{\kappa} < c_s < \overline{\kappa}$), so that the processing costs exceed the benefits of knowing s when $\theta = \theta^L$, but the benefits exceed the costs when $\theta = \theta^H$, the conditional strategy, $(\theta, (s, \not s))$, is optimal. By not processing s when $\theta = \theta^L$, the agent avoids incurring costs that exceed the benefits. The agent processes s only when $\theta = \theta^H$, i.e., when the benefits exceed the costs.

Overall, Proposition 2 reveals that the allocation of attention is context-dependent; dependent on the processing costs. The thresholds $\hat{\kappa}$, $\underline{\kappa}$, and $\overline{\kappa}$ that determine the tipping points are also context-dependent. $\underline{\kappa}$ and $\overline{\kappa}$ are dependent on c_θ and the values of $\theta^H,\ \theta^L,\ s^+,\ s^-,\ p_s$, and p_θ. $\hat{\kappa}$ and $\underline{c}_\theta$ depend also on the values of $\theta^H,\ \theta^L,\ s^+,\ s^-,\ p_s$, and p_θ.

Lemma 1 *Behavior of $\underline{c}_\theta$, $\hat{\kappa}$, $\underline{\kappa}$, and $\overline{\kappa}$:*

i) $\underline{c}_\theta$ increases with θ^H. If $E[s] \leq 0$, $\underline{c}_\theta$ increases with s^+ and p_s. $\underline{c}_\theta$ decreases with θ^L. If $E[s] > 0$, $\underline{c}_\theta$ decreases with s^- and p_s. If $p_\theta < 0.5$, $\underline{c}_\theta$ increases with p_θ. If $p_\theta > 0.5$, $\underline{c}_\theta$ decreases with p_θ.

ii) $\hat{\kappa}$ increases with θ^H, θ^L, and p_θ. If $E[s] \leq 0$, $\hat{\kappa}$ also increases with s^+ and p_s. If $E[s] > 0$, $\hat{\kappa}$ decreases with s^- and p_s.

iii) $\underline{\kappa}$ increases with θ^L, p_θ, and c_θ. If $E[s] \leq 0$, $\underline{\kappa}$ increases also with s^+ and p_s. If $E[s] > 0$, $\underline{\kappa}$ decreases with s^- and p_s.

iv) $\overline{\kappa}$ increases with θ^H and p_θ. If $E[s] \leq 0$, $\overline{\kappa}$ also increases with s^+ and p_s. $\overline{\kappa}$ decreases with c_θ. If $E[s] > 0$, $\overline{\kappa}$ decreases with s^- and p_s.

An increase of $\hat{\kappa}$, $\underline{\kappa}$, and $\overline{\kappa}$ translates into a shift to the right in Figure 3, a decrease into a shift to the left. Thus as $\hat{\kappa}$ and $\underline{\kappa}$ increase, the area in which $(s, (\not\theta, \not\theta))$ is optimal increases. As $\overline{\kappa}$ increases, the area in which $(\theta, (s, \not s))$ is optimal increases. Consequently, the area in which a strategy is optimal, as illustrated in Figure 3, also depends on these variable values.

A higher p_θ puts more probability on the high-stake situations, so in expectations more utility is at stake. Thus the benefits from knowing s are higher and processing s becomes optimal for higher processing costs. This translates into a shift of $\hat{\kappa}$, $\underline{\kappa}$, and $\overline{\kappa}$ to the right. As θ scales the utility difference between options 1 and 2, if θ^L and/or θ^H increase, more utility is at stake. Thus costs for processing s may also be higher, because the trade-off is still in favor of knowing the realization of the variable s. Thus $\hat{\kappa}$ and $\overline{\kappa}$ increase in θ^H and $\hat{\kappa}$ and $\underline{\kappa}$ increase in θ^L.

As a higher s^+ or a higher (in absolute terms) s^- scale up the utility difference for θ^L and for θ^H, the benefits of knowing s increase. Thus processing s is optimal also for higher costs. This translates into a shift of $\hat{\kappa}$, $\underline{\kappa}$, and $\overline{\kappa}$ to the right. If p_s increases, $\hat{\kappa}$, $\underline{\kappa}$, and $\overline{\kappa}$ increase for $E[s] \leq 0$, but decrease for $E[s] > 0$. The intuition is as follows: The higher the uncertainty about whether option 1 or option 2 is better—that is the more uncertain it is whether $s = s^+$ or $s = s^-$—the higher the benefits of knowing s. Consequently, $\hat{\kappa}$, $\underline{\kappa}$, and $\overline{\kappa}$ shift to the right. As $p_s \longrightarrow 0$ or $p_s \longrightarrow 1$, uncertainty vanishes. The agent incurs almost no loss by choosing given expectations. Consequently, $\hat{\kappa}$, $\underline{\kappa}$, and $\overline{\kappa}$ shift to the left. So the maximum shift to the right occurs as the expected utility of options 1 and 2 converge.

As c_θ increases, processing θ becomes more expensive. As $(\theta, (s, \not s))$ is the only strategy in which θ is processed, the area in which $(\theta, (s, \not s))$ is optimal decreases. With increasing c_θ it becomes optimal to either process s directly or to process nothing.

$\underline{c}_\theta$ determines the threshold on the costs for processing θ below which the conditional strategy exists. So whenever the utility difference between options 1 and 2 increases, e.g., θ^H increases or θ^L decreases, it becomes optimal to condition on the importance. Thus $\underline{c}_\theta$ increases. In addition, the higher the certainty about which θ occurs, the less relevant it is to condition on θ, instead it makes sense to process s always or never. Thus as $p_\theta \longrightarrow 0$ or $p_\theta \longrightarrow 1$, $\underline{c}_\theta$ decreases. Or the other way round, as $p_\theta \longrightarrow 0.5$, $\underline{c}_\theta$ increases. Similar considerations regarding p_s and s as discussed for $\hat{\kappa}$, $\underline{\kappa}$, and $\overline{\kappa}$ hold for $\underline{c}_\theta$.

The model allows an analysis of the link between processing costs and decision-making. Processing costs have three major implications: First, processing costs determine choice quality. Second, lowering processing costs can crowd out information. Third, processing costs can induce choice reversals. In the following I elaborate on these three implications and, in addition, comment on the link between the resulting attention allocation and the psychological concepts of top-down and bottom-up attention.

As the processing costs determine which strategy is optimal, the processing costs determine whether the agent chooses between options 1 and 2 given information or expectations. Only if the agent processes s does she choose the utility-maximizing option with certainty. Thus with strategy $(s, (\not\emptyset, \not\emptyset))$, she always chooses the utility-maximizing option. With strategy $(\theta, (s, \not s))$, if $\theta = \theta^H$, i.e., with probability p_θ, the agent also chooses the utility-maximizing option. But if $\theta = \theta^L$, it depends on chance: The agent chooses an option given the expected value of s. So if $E[s] > 0$, she chooses option 1, but with probability $1 - p_s$, $s = s^-$ and the agent incurs a loss. If $E[s] < 0$, she chooses option 2 but with probability p_s, $s = s^+$ in which case option 1 would have been optimal. If $E[s] = 0$, the agent randomizes and might by chance choose the wrong option. The probability that the agent chooses the wrong option and incurs a loss with strategy $(\theta, (s, \not{s}))$ is thus

$$\begin{cases} (1-p_\theta)p_s & \text{if } E[s] < 0 \\ (1-p_\theta)(0.5p_s + 0.5(1-p_s)) = (1-p_\theta)0.5 & \text{if } E[s] = 0 \\ (1-p_\theta)(1-p_s) & \text{if } E[s] > 0. \end{cases}$$

With strategy $(\not{\theta}, \not{s})$, the agent always decides given expectations. The probability that the agent chooses the wrong option is

$$\begin{cases} p_s & \text{if } E[s] < 0 \\ 0.5 & \text{if } E[s] = 0 \\ (1-p_s) & \text{if } E[s] > 0. \end{cases}$$

So the agent does not always choose the option with the highest actual utility. The probability with which she incurs a loss depends on the chosen attention strategy and the expected value of s. As the selection of the attention strategy depends on the processing costs, the quality of the choice ultimately also depends on the processing costs.

The model also predicts that information can be crowded out by lowering processing costs. Figure 4 illustrates this. At point A the agent uses strategy $(s, (\not{\theta}, \not{\theta}))$. So, the agent processes s and thus always chooses the utility-maximizing option. With a lower c_θ at point B, the agent uses strategy $(\theta, (s, \not{s}))$. Although the costs for processing s are the same, the probability of processing s decreased from 1 to p_θ. The lower c_θ induces the processing of θ at the expense of s. In addition, the change in processing costs influences the choice quality. A further example shows that decreasing the costs of processing s crowds out the processing of θ. Starting from point B and decreasing c_s, leads to point C. Here the agent uses strategy $(s, (\not{\theta}, \not{\theta}))$. Thus, by decreasing the costs of processing s, s is processed more often (always). But, although c_θ has not changed, θ is processed no more. Therefore, by decreasing the costs of processing a variable, the processing of that variable can be induced at the expense of no longer or less often processing other information. This change in processing costs also impacts choice quality.

This example also illustrates that the model exhibits choice reversals. Table 2 summarizes all possible situations that can occur and the corresponding choices for each attention strategy. If s is in line with expectations, i.e., $s = s^-$ and $E[s] \leq 0$ or $s = s^+$ and $E[s] > 0$, choice reversals do not occur. Each level of information leads to the same choice. Choice reversals occur if s is not in line with expectations and the level of information changes. For example, if $s = s^+$, $\theta = \theta^H$, and $E[s] < 0$, under inattention the agent chooses option 2. If

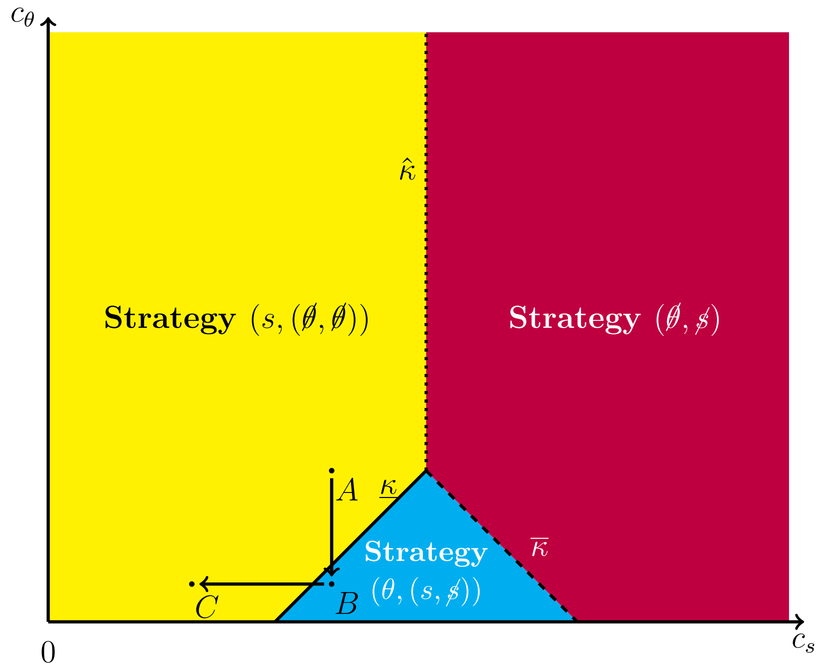


Figure 4: Illustration of crowding out.

c_s drops such that the agent selects $(s, (\not\theta, \not\theta))$, the agent processes s and knows that option 1 is better. Thus she chooses option 1 as a consequence of more information. Thus in the model choice reversals occur as a consequence of changes in information. Consequently, the model offers a stylized explanation for changes in observed behavior that occur as salience of information increases (e.g., Chetty, Looney, and Kroft, 2009). The increase in salience leads to processing of information that was previously not processed (such as tax-inclusive prices).

		s^+, θ^H	s^-, θ^H	s^+, θ^L	s^-, θ^L
	$(\not\theta, \not s)$	option 2	option 2	option 2	option 2
$E[s] \leq 0$	$(s, (\not\theta, \not\theta))$	option 1	option 2	option 1	option 2
	$(\theta, (s, \not s))$	option 1	option 2	option 2	option 2
	$(\not\theta, \not s)$	option 1	option 1	option 1	option 1
$E[s] > 0$	$(s, (\not\theta, \not\theta))$	option 1	option 2	option 1	option 2
	$(\theta, (s, \not s))$	option 1	option 2	option 1	option 1

Table 2: Agent's choices for each attention strategy and each possible situation. The gray cells highlight the cases, when due to changing processing costs, choice reversals may occur.

Proposition 2 demonstrates that two attention strategies prevail in addition to inattention. I argue that these two attention strategies show similarities to the two attention mechanisms *bottom-up* and *top-down* attention described in the psychological literature.[8] Bottom-up attention refers to information selection which is driven by exogenous factors. The salience of the stimuli draws attention automatically. In contrast, top-down attention is driven by cognitive processes and goals, i.e., it is endogenous, and requires more effort. In comparison bottom-up attention is faster.

Strategy $(\theta, (s, \not s))$, the conditional allocation of attention, shares characteristics with top-down attention. Being conditional makes $(\theta, (s, \not s))$ more complex than a direct selection. In addition, as $(\theta, (s, \not s))$ requires a decision about what implications the processing of θ has on the processing of s, $(\theta, (s, \not s))$ is more endogenous. Furthermore, the processing of two variables (instead of one variable with $(s, (\not\theta, \not\theta)))$ makes $(\theta, (s, \not s))$ slower. This is true especially if time costs are part of processing costs. The processing induced by $(\theta, (s, \not s))$ includes higher processing costs if s is processed (i.e., if $\theta = \theta^H$).

Strategy $(s, (\not\theta, \not\theta))$ shares characteristics with bottom-up attention. The processing of s is faster with $(s, (\not\theta, \not\theta))$ than with $(\theta, (s, \not s))$ because s is processed directly without detour via θ. Bottom-up attention is induced because of the salience of the information which automatically draws attention. In comparison, $(s, (\not\theta, \not\theta))$ is used when the costs for processing s are (comparatively) low. The costs of the good-specific s are more relevant for the selection of $(s, (\not\theta, \not\theta))$ than the costs of processing of the not necessarily good-specific θ.[9] So the ease with which s can be processed induces the use of strategy $(s, (\not\theta, \not\theta))$. Moreover, only s is directly processed. This is in line with research revealing that reward associated with a stimuli induces processing similar to pure salience (Pessoa, 2014).[10] As s determines which option is better, only s is associated with receiving rewards. The variable θ determines the level of reward, but is not decisive for whether a reward is gained. Therefore the agent has no incentive to process θ directly without the objective to condition the processing of s on θ (Proposition 1 and 2). Thus the model accounts for different types of information and the costs alone do not determine whether the information are processed.

That rational considerations of the agent in the model result in the use of two attention strategies that share similarities with attention mechanisms reported in psychology is valuable. This finding offers a rational explanation for the existence of top-down and bottom-up

[8]See, for example, Nobre and Kastner (2014b), especially also Nobre and Mesulam (2014) for a brief overview. The notation in the literature varies. Notations include, for instance, endogenous/voluntary/effortful and exogenous/reflexive/automatic (Nobre and Mesulam, 2014). See also *system 1* and *system 2* (Kahneman, 2003, 2011). I will exclusively refer to the concepts as top-down and bottom-up attention.

[9]Only when $\underline{c}_\theta > c_\theta$ is c_θ relevant, but then only marginally.

[10]See, for example, also Pashler, Johnston, and Ruthruff (2001) who argue that top-down goals actually influence what captures attention bottom-up. This would be in line with the finding that only s is directly processed. The goal of the agent is to choose the option with the highest utility. For this goal, processing s is relevant.

attention. In addition, this finding—that the resulting attention allocation shares similarities with concepts reported in the psychological literature—shows that the results of the model are not detached from reality and thus strengthens the argumentation.

Changing the perspective, by assuming that a link between $(s, (\not\theta, \not\theta))$ and bottom-up and $(\theta, (s, \not s))$ and top-down attention exists, the model offers predictions about when bottom-up and top-down attention and inattention are used in decision-making: A decision-maker uses bottom-up attention in a low cost environment, when benefits exceed the costs, e.g., when the decisive variable is salient. If costs are intermediate, the decision-maker uses top-down attention and in a high cost environment she uses inattention. Relevant for the selection are thus the processing costs.

5 Strategic Considerations of Firms: Is Exploiting Consumers Optimal?

This section takes the attention allocation derived in the previous section as a premise of how consumers allocate attention and examines the strategic considerations of a firm. Assume a competitive market with $n > 1$ firms that produce a good. The price of the good of all firms is identical and fixed at ρ. The good is identical, except for quality, which can be high or low, and the processing costs the consumers have for processing the quality. Each firm $i \in \{1, 2, ..., n\}$ can set the quality of the good s_i[11] and the consumers' costs for processing the quality, $c_s^i \in (0, \infty)$, individually. Setting $c_s^i \in (0, \infty)$ corresponds to shrouding or highlighting the quality. For each $c_s^i \in (0, \infty)$, the firm has consideration costs (i.e., why it sets this value) and implementation costs. On the one hand, setting c_s^i low the firm has costs for advertising the quality or printing it prominently on the packaging. On the other hand, setting c_s^i high, so that quality is shrouded, the firm has costs for distracting the consumer from the quality by advertising something else, or printing other information prominently on the packaging. Therefore, I assume that the costs of setting c_s^i are zero for all firms and all values of c_s^i.

Assume a unit mass of identical consumers. Consumers are randomly distributed across firms and then decide between buying (option 1 with $u_1 = \theta s_i$) and not buying (option 2 with $u_2 = 0$). The firm thus operates as a local monopoly but consumers' expectations are shaped by the whole market. The probability that a consumer looks at the good of firm

[11] Technically, s_i represents whether a consumer prefers buying the good of firm i to not buying the good. Thus the variable determines whether the quality exceeds the price. If the quality is high, the quality exceeds the price ($s_i = s^+$). If the quality is low, the price exceeds the quality ($s_i = s^-$). In the following, as the price is fix for all firms, I will refer to s_i as quality.

$i \in \{1, 2, ..., n\}$ depends on the number of firms: $r_i = \frac{1}{n} \forall i$. As the consumer population is normalized to 1, r_i specifies the fraction of consumers looking at firm i's good.[12] The utility consumers derive from the good of firm i (option 1) depends on the quality s_i, which firm i sets, and the state of the world $\theta \in \{\theta^L, \theta^H\}$, which is exogenously given. Consumers allocate attention rationally as specified in section 3 to process θ and s_i.

Firms can neither influence θ, nor the costs consumers have for processing θ. Firms choose quality and perception costs of quality to maximize profits. Therefore, firm i's objective is

$$\max_{s_i, c_s^i} \{\Pi_i = (\rho - K(s_i))x_i^D\},$$

where x_i^D represents the demand for the good of firm i and $K(s_i)$ the production costs. For simplicity, assume that the production costs are identical for all firms. Without loss of generality, we can set production costs to 0 and only impose costs if the firm produces high quality, i.e., let $K(s^+) \equiv k > 0$ and $K(s^-) \equiv 0$. The costs are only incurred upon sale. The demand for the firm of good i, x_i^D, depends on

(1) r_i, the fraction of consumers allocated to firm i. The fewer firms are in the market, the higher the fraction and, thus, the higher—potentially—the demand.

(2) c_s and c_θ. The processing costs determine which attention strategy the consumers use. In turn, the attention strategy determines whether the consumers choose between buying and not buying given the true quality or the expected quality of the good.

(3) s_i, the quality of the good. If consumers choose an attention strategy with which they process the quality, the quality determines whether they buy. If $s_i = s^+$ they buy and if $s_i = s^-$ they do not buy the good.

(4) $E[s]$, the expected quality. If consumers choose an attention strategy with which they do not process the quality, the expected quality of the good determines whether they buy. If $E[s] > 0$ they buy, if $E[s] = 0$ they randomize and if $E[s] < 0$ they do not buy the good.

(1) is an exogenous factor for the firm. Nevertheless, the firm can partially influence which attention strategy its consumers use (2) by setting c_s^i, i.e., by highlighting or shrouding information on quality. The firm can always induce attention strategies $(\not\theta, \not s)$ and $(s, (\not\theta, \not\theta))$, because regions where attention strategies $(\not\theta, \not s)$ and $(s, (\not\theta, \not\theta))$ are optimal can always be reached by setting c_s^i accordingly (see Figure 3).[13] But, if $c_\theta > \underline{c}_\theta$, the firm cannot induce

[12] An alternative interpretation assumes a representative consumer and r_i the probability with which the representative consumer visits firm i.

[13] Exception: If $p_s = 0$ and $E[s] \leq 0$, or if $p_s = 1$ and $E[s] > 0$. Then agents are always inattentive.

strategy $(\theta, (s, \not s))$, because it has no control over c_θ (see also Figure 3). In addition, the firm can fully influence the quality it produces (3) and thus influence the expected quality in the market (4). If the firm sets the processing costs such that consumers process the quality, the firm can influence whether its product is bought by producing high quality. But if consumers do not process the quality, the consumers choose given expectation. The firm is only able to influence the decision of the consumer, if it is able to tip the expected quality, $E[s]$. For example, if $E[s] < 0$, consumers do not buy. If the firm is able to tip $E[s] < 0$ to $E[s] = 0$ by changing its quality from $s_i = s^-$ to $s_i = s^+$ (firm i is pivotal), consumers randomize. However, the firm can influence whether its quality is processed.

Remark 1 *If $E[s] < 0$ and $\rho > k$, a non-pivotal, profit-maximizing firm always has an incentive to produce high quality and highlight this. If $E[s] > 0$, a non-pivotal profit-maximizing firm always has an incentive to produce low quality and shroud this.*

The proof proceeds by evaluating how different attention strategies and the quality influence demand (and profit). Overall Remark 1 shows that the firm has an incentive to produce quality contrary to the quality consumers expect. If consumers expect low quality, they do not buy unless the firm produces high quality and they process the quality. Thus to make a sale, the firm has to produce high quality and highlight the quality such that consumers process the quality. Consumers profit from the firm producing high quality and highlighting this so that they buy a high quality good which they prefer to not owning the good. If the expected quality is positive, the firm maximizes profit by producing low quality. If consumers expect high quality, they buy even if they do not process the quality. The firm can exploit this by producing low quality—i.e., avoiding high production costs—and still making a sale.[14] Thus consumers buy expecting a high quality good but receive a low quality good, although they prefer not buying to low quality. Their inattention, which was induced by the firm on purpose, is exploited. A firm never attempts to induce the conditional attention allocation. This is intuitive. If a firm can influence the likelihood with which consumers process the quality, the firm prefers that consumers always process the quality if it is high, and never if it is low. So that firms are not always able to induce the conditional attention strategy is irrelevant.

It is relevant though that the firm is not pivotal. If the firm is pivotal in the sense that it changes the consumers' expectations by their quality choice, the firm's incentives are different. This is illustrates in Proposition 3.

Assumption 1 *The number of firms n is such that $E[s] = 0$ is possible.*

[14] If $E[s] = 0$, the strategy of the firm depends on k. However, firms also decide between low quality with shrouding and high quality with highlighting.

Assumption 1 is necessary for the equilibrium in Proposition 3 to exist. Assumption 1 implies, for example, that if $s^+ = 1$ and $s^- = -1$, n has to be even. Because for $E[s] = 0$, the fraction of firms producing high quality has to be $p_s = \frac{1}{2}$. Thus with a finite number of firms, n has to be even.

Proposition 3 *Nash equilibria.*

If $k > \rho$, there exists a continuum of Nash equilibria in which all firms produce low quality.

If $k < \rho < 2k$, there exists a continuum of Nash equilibria in which

i) producers who produce low quality shroud quality

ii) producers who produce high quality highlight quality

and the fraction of high-quality producers p_s is such that $E[s] = 0$.

If $\rho > 2k$, there exists a continuum of Nash equilibria in which

i) producers who produce low quality shroud quality

ii) producers who produce high quality highlight quality

and the fraction of high-quality producers is $p'_s = min\{p_s | E[s] > 0\}$.

If $k > \rho$, a profit-maximizing firm never produces high quality. Consumers reasonable expect this and do not buy. Thus profit is 0 and consumers' utility is 0. If $k < \rho < 2k$, to produce high quality is profitable. By Proposition 1, if $E[s] < 0$, the optimal strategy is to produce high quality and if $E[s] > 0$ the optimal strategy is to produce low quality. So whenever the market is out of equilibrium, a trend towards $E[s] = 0$ exists. If $E[s] = 0$, no firm has an incentive to deviate, because by changing their quality, the distribution changes and thus $E[s] \neq 0$. As the strategies of firm i consist of setting c_s^i in addition to s_i, shrouding or highlighting simply stands for $c_s^i > \overline{\kappa}$ or $c_s^i > \hat{\kappa}$ and $c_s^i < \underline{\kappa}$ or $c_s^i < \hat{\kappa}$. But c_s does not have to take on a specific value. This is the reason why a continuum of equilibria exists. Thus in equilibrium on average consumers receive the same utility by buying (option 1) than by not buying (option 2). If $\rho > 2k$, a firm makes higher profits by producing high quality when $E[s] > 0$ than by producing low quality when $E[s] = 0$. Therefore, although the firm prefers to produce low quality when $E[s] > 0$, the pivotal firm has no incentive to deviate to producing low quality.

Overall the costs for processing quality are indicative about the quality of the good. However, consumers do not incorporate this in their considerations. They are rational in their attention allocation by maximizing expected utility, but naïve in the sense of ignoring firms strategic considerations. In addition, in all equilibria of Proposition 3 if $k < \rho$, some consumers are exploited.

6 Conclusion

This paper proposes a model of rational allocation of attention in decision-making. Exploring the resulting attention allocation reveals that the selection of an attention strategy, and consequently the quality of the choice between the two options, are highly context-dependent. The always fully informed agent, as argued by standard theory, only shows up in the limiting case. Generally, agents do not always gather all available information. Therefore, agents are not always informed about which option they prefer. This result is in line with economic findings that decision-makers do not always fully process all information and that salience is relevant to induce processing (Chetty, Looney, and Kroft, 2009; Finkelstein, 2009). In my model, as the processing costs, interpretable as the level of salience, change, the optimal attention strategy changes. Thresholds mark the tipping points. For example, as it becomes easier to process the relevant information (costs for processing s decrease) and costs move beyond the threshold, the agent moves from inattention to processing s and thus from choosing given expectations to choosing given information. Consequently, the consumption behavior can change and average quality of choice improve. Changes in consumption behavior are thus a result of using different attention strategies.

In the resulting attention allocation, two attention strategies prevail (next to inattention) that share similarities with top-down and bottom-up attention, concepts from the psychological literature. The model thus offers an explanation for the existence of top-down and bottom-up attention; as the result of maximizing utility. The resulting attention allocation becomes the premise of how consumers allocate attention in section 5, which investigates the strategic considerations of firms. That the attention strategies share similarities with observed mechanism is advantageous as it provides evidence for the premise of how consumers allocate attention and thus strengthens the soundness of the argument.

Exploring the strategic considerations of a firm shows that a firm has an incentive to shroud information on quality only in combination with low quality.[15] Overall, a firm has an incentive to produce quality contrary to consumers' expectations. If consumers expect low quality, the firm has an incentive to produce high quality and highlight this. This is intuitive:

[15] Brown, Hossain, and Morgan (2010) provide evidence from online auctions that "(1) shrouding affects revenues—for low shipping charges, a seller is better off disclosing; and (2) increasing shipping charges boosts revenues when shipping charges are shrouded."(Brown, Hossain, and Morgan, 2010, p. 859f.) This is in line with the result here that shrouding low quality and highlighting high quality is optimal.

A consumer who expects low quality does not buy unless the firm shows that it sells high quality. In this case consumers benefit. But if consumers expect high quality, the firm has an incentive to produce low quality and shroud this. The consumer then buys expecting high quality but receives low quality, i.e., the firm exploits the consumer. An equilibrium exists in which firms producing low quality shroud quality and firms producing high quality highlight quality. The distribution of high and low quality is such that $E[s] = 0$, i.e., $p_s = \frac{-s^-}{s^+ - s^-}$. For example, if $s^+ = 1$ and $s^- = -1$, then $p_s = 0.5$. In equilibrium, the expected utility from option 1 equals the utility of option 2.

Understanding the strategic considerations of firms is essential for assessing instruments to prevent the exploitation of consumers. Two instruments are conceivable in the model. Enforcing transparency (making it easy to process the relevant information) or setting quality standards. Enforcing transparency translates to enforcing a limit on processing costs at a level such that agents always process the relevant information, i.e., when strategy $(s, (\not\theta, \not\theta))$ is optimal. In reality, all firms may, for example, have to print a seal of quality or label on the product so that the consumers can inform themselves easily about the quality. The seal of quality has to be extremely easy to process. A seal that makes inferring the quality too complex may even increase the costs of the consumers to assess the quality. Setting a quality standard regulates the distribution of s in the market. If $p_s = 1$, decisions given expectations lead to choosing the best option; attention is not needed. Therefore, the policy options are to frame choices to make people pay attention or to make it irrelevant for people to pay attention. From a welfare perspective, ceteris paribus, for the consumers the second option is better. Because paying attention to s leads to an expected utility of $E_{(s,(\not\theta,\not\theta))} = E[\theta]p_s s^+ - c_s$, whereas not paying attention when $p_s = 1$ gives an expected utility of $E_{(\not\theta,\not s)} = E[\theta]s^+$. $E_{(s,(\not\theta,\not\theta))}$ can never be higher than $E_{(\not\theta,\not s)} = E[\theta]s^+$ and is only as high if $p_s = 1$ and $c_s = 0$.[16]

The model offers solutions for policy makers other than transparency and quality standards. In situations when agents are faced with a yes (option 1) or no (option 2) decision, e.g., getting a vaccination or not getting a vaccination,[17] educating agents proves very costly, and the social planner has a clear preference about which options agents ought to choose, the model supports the possibility of establishing transfers or non-monetary incentives. If processing costs are high, agents choose given expectations. If the expected value is negative they choose option 2, if it is positive they choose option 1. The social planner can influence this decision by offering transfers (or non-monetary incentives) t. This might tip the decision.

[16]However, this assumes that for all agents the choice variable signifies the same thing. So in a scenario with heterogeneous s, e.g., when for some agents the quality of a laptop is determined the weight and for others by battery life, transparency might be the better option.

[17]See, for instance, Banerjee, Duflo, Glennerster, and Kothari (2010) for an immunization study with incentives.

If $E[u_1] < 0$, but the social planner would like agents to choose option 1, the social planner has to set t such that $E[u_1] + t > 0$. Agents then have a positive utility from option 1, although they expect the option itself to be worse than option 2. The transfer changes the total expected utility and so induces behavioral change without having to educate the agent.

Some issues are beyond the scope of this paper and may serve as ideas for future research. For example, in section 5, I assume that agents allocate attention rationally. Yet, consumers do not infer from the shrouding of firms that the firm produces low quality.[18] Additionally, I assume that the quality signaling of firms is believable. The model also implicitly assumes that agents are aware of the costs c_s and c_θ, before choosing an attention strategy. This suggests either direct observation or extremely fast learning. Nevertheless, slower learning, modeled via replicator dynamics,[19] leads to similar results. In the areas where $(\not\theta, \not s)$, $(s, (\not\theta, \not\theta))$, and $(\theta, (s, \not s))$ are strictly optimal, respectively the other strategies vanish. But learning as modeled via replicator dynamics require repeatedly facing the same decision.

The model includes only two variables to which the agent can allocate attention. Therefore, the model does not allow for clear predictions about options with multiple dimensions, unless two options are mostly identical and differ only in one dimension. For example, two notebooks are identical, but differ in weight. In the model option 1 captures this difference: θ is the importance the agent assigns to the weight dimension. For some agents weight is extremely important ($\theta = \theta^H$), for instance, because she travels frequently, for others weight matters little compared to other dimensions ($\theta = \theta^L$). s illustrates which notebook is less heavy.

Although the model specifies a (consumption) decision between two options, the model can be applied to a range of situations. First, options do not have to be goods. They can represent parties, jobs, or any other options between which decisions are made. Option 2 can then still represent the outside option. But option 2 can also be a default (such as a favorite brand or current job) against which the agent compares a new option (new brand or new job). Or, options 1 and 2 are simply any two options from a set of many options. In both cases s is a relative variable—capturing which option is better—and θ can additionally be interpreted as a measure for how different the options are. For example, if a pivotal voter has to decide between two parties, θ measures how different the policy platforms are, and s which policy program is more in line with the voter's preferences.

Second, the setup can also be extended to describe situations where multiple options

[18] Still this might be a reasonable assumption, because under the load of information a consumer is confronted with, she might not notice that the firm does not disclose the relevant information but instead distracts her. Especially considering that this model only focuses on one decision, while a consumer—for example, in the supermarket—has to make a multitude of decisions.

[19] See Weibull (1995) for textbook treatment.

exist. Option 2 represents the market average, whereas option 1 is a randomly picked good. s represents whether a good's quality exceeds the market average. θ can be interpreted as the diversity in the market. Markets exist where quality diverges little $\theta = \theta^L$, e.g., in the paperclip market, where quality differences are marginal. But there also exist markets where differences are enormous $\theta = \theta^H$—a lot of high and a lot of low quality products exist, e.g., in the clothes market.

As an alternative to capture situations with multiple options, options are compared pairwise. Starting with any option, this option represents the default, option 2 (utility normalized to 0). Then, the agent picks another option at random and compares it to the default. If the agent perceives it to be better, it becomes the new default, otherwise the new option is discarded and the old default is kept. Afterwards, the agent picks a new option at random and compares it to the default. This pairwise comparison continues until the agent has considered all options. The agent will choose the last default.[20]

Overall, the model can capture a variety of situations and explains interactions between processing costs and choice. In addition, the link to psychological literature supports the model's attention allocation and the model offers a explanation for the existence of top-down and bottom-up attention.

[20]This interpretation needs to include learning of θ. As θ is not good-specific, if the agent processes it once, the only costs incurred in the remainder are recall costs. So costs for processing θ should decrease over time.

References

Banerjee, A. V., E. Duflo, R. Glennerster, and D. Kothari (2010): "Improving Immunization Coverage in Rural India: A Clustered Randomized Controlled Evaluation of immunization Campaigns with and without Incentives," *British Medical Journal*, 340.

Brown, J., T. Hossain, and J. Morgan (2010): "Shrouded Attributes and Information Suppression: Evidence from the Field," *Quarterly Journal of Economics*, 125, 859–876.

Chetty, R., A. Looney, and K. Kroft (2009): "Salience and Taxation: Theory and Evidence," *American Economic Review*, 99, 1145–1177.

Eliaz, K., and R. Spiegler (2011): "On the strategic use of attention grabbers," *Theoretical Economics*, 6, 127–155.

Finkelstein, A. (2009): "E-ztax: Tax Salience and Tax Rates," *Quarterly Journal of Economics*, 124, 969–1010.

Gabaix, X. (2014): "A Sparsity-Based Model of Bounded Rationality," *Quarterly Journal of Economics*, 129, 1661–1710.

Gabaix, X., and D. Laibson (2006): "Shrouded Attributes, Consumer Myopia, and Information Suppression in Competitive Markets," *Quarterly Journal of Economics*, 121, 505–540.

Heidhues, P., B. Köeszegi, and T. Murooka (2015): "Exploitative Innovation," *American Economic Journal: Microeconomics*, 8, 1–23.

——— (2016): "Inferior Products and Profitable Deception," *Review of Economic Studies*, forthcoming.

Hossain, T., and J. Morgan (2006): "...Plus Shipping and Handling: Revenue (Non) Equivalence in Field Experiments on eBay," *Advances in Economic Analysis & Policy*, 6, Article 3.

Kahneman, D. (2003): "Maps of Bounded Rationality: Psychology for Behavioral Economics," *American Economic Review*, 93, 1449–1475.

——— (2011): *Thinking, fast and slow*. Penguin Books, London.

Manzini, P., and M. Mariotti (2014): "Stochastic Choice and Consideration Sets," *Econometrica*, 82, 1153–1176.

Nobre, A. C., and S. Kastner (2014a): *Attention: Time Capsule 2013.* in: Nobre, A.C. and S. Kastner: The Oxford Handbook of Attention, Oxford University Press, Oxford.

——— (2014b): *The Oxford Handbook of Attention.* Oxford University Press, Oxford.

Nobre, A. C., and M. Mesulam (2014): *Large-Scale Networks for Attentional Biases.* in: Nobre, A.C. and S. Kastner: The Oxford Handbook of Attention, Oxford University Press, Oxford.

Pashler, H., J. C. Johnston, and E. Ruthruff (2001): "Attention ane Performance," *Annual Review of Psychology*, 52, 629–651.

Pessoa, L. (2014): *Attention, Motivation, and Emotion.* in: Nobre, A.C. and S. Kastner: The Oxford Handbook of Attention, Oxford University Press, Oxford.

Sims, C. (2003): "Implications of rational inattention," *Journal of Monetary Economics*, 50, 665–690.

Weibull, J. W. (1995): *Evolutionary Game Theory.* MIT Press, Cambridge, Mass.

A Expected Utilities of Attention Strategies

The expected utility of an attention strategy corresponds to the expected utility of the chosen option and the costs for processing incurred with that strategy. The utility of option 1 is $u_1 = \theta s$ with $0 < \theta^L < \theta^H$ and $s^- < 0 < s^+$. As θ and s are independently distributed, the expected utility of option 1 is $E[u_1] = E[\theta]E[s]$. The utility of option 2 is $u_2 = 0$. Agents know the utility of option 2: $E[u_2] = u_2 = 0$. The costs for processing are always strictly positive: $c_\theta > 0$ and $c_s > 0$.

Attention strategy $(s, (\not\emptyset, \not\emptyset))$**:** The agent processes only s. Thus she has costs for processing s. As she always knows the value of s, she knows whether option 1 yields a higher or lower utility than option 2. If $s = s^+$, she chooses option 1, because $E[u_1] = E[\theta]s^+ > 0 = u_2$, and if $s = s^-$, she chooses option 2, because $E[u_1] = E[\theta]s^- < 0 = u_2$. The expected utility associated with strategy $(s, (\not\emptyset, \not\emptyset))$ is thus:

$$E_{(s,(\not\emptyset,\not\emptyset))} = p_s E[u_1] + (1 - p_s)u_2 - c_s = E[\theta]p_s s^+ - c_s.$$

Attention Strategy $(s, (\not\emptyset, \theta))$**:** The agent always processes s, but processes θ only if $s = s^-$. If $s = s^+$, she chooses option 1 and if $s = s^-$, she chooses option 2. As the agent processes θ only if $s = s^-$, she incurs costs for processing θ with the probability that $s = s^-$, i.e., $1 - p_s$. Nevertheless, she always has costs for processing s, c_s:

$$\begin{aligned} E_{(s,(\not\emptyset,\theta))} &= p_s u_1 + (1 - p_s)u_2 - c_s - (1 - p_s)c_\theta \\ &= E[\theta]p_s s^+ - c_s - (1 - p_s)c_\theta. \end{aligned}$$

Attention Strategy $(s, (\theta, \not\emptyset))$**:** The agent always processes s, but processes θ only if $s = s^+$. If $s = s^+$, she chooses option 1 and if $s = s^-$, she chooses option 2. As the agent processes θ only if $s = s^+$, she incurs costs for processing θ with the probability that $s = s^+$, i.e., p_s. Nevertheless, she always has costs for processing s, c_s:

$$\begin{aligned} E_{(s,(\theta,\not\emptyset))} &= p_s u_1 + (1 - p_s)u_2 - c_s - p_s c_\theta \\ &= E[\theta]p_s s^+ - c_s - p_s c_\theta. \end{aligned}$$

Attention Strategy (θ, s)**:** The agent processes θ and s. Thus she is able to calculate u_1 precisely and always choose the utility-maximizing option. If $s = s^+$, she chooses option 1, because either $u_1 = \theta^H s^+ > 0 = u_2$ (if $\theta = \theta^H$) or $u_1 = \theta^L s^+ > 0 = u_2$ (if $\theta = \theta^L$). If $s = s^-$, she chooses option 2, because either $u_1 = \theta^H s^- < 0 = u_2$ (if $\theta = \theta^H$) or $u_1 = \theta^L s^- < 0 = u_2$ (if $\theta = \theta^L$). Nevertheless, she also always has costs for processing θ and s. Thus, the expected utility of strategy (θ, s) is:

$$\begin{aligned}
E_{(\theta,s)} &= p_\theta \left(p_s u_1 + (1-p_s)u_2\right) + (1-p_\theta)\left(p_s u_1 + (1-p_s)u_2\right) - c_\theta - c_s \\
&= p_\theta p_s \theta^H s^+ + (1-p_\theta) p_s \theta^L s^+ - c_\theta - c_s \\
&= E[\theta] p_s s^+ - c_\theta - c_s.
\end{aligned}$$

Attention strategy $(\not\theta, \not s)$**:** the agent processes neither θ nor s. Thus she has no perception costs, but has to decide between options 1 and 2 given her expectations about s. If $E[s] < 0$, the agent expects that $E[u_1] = E[\theta]E[s] < 0 = u_2$ and therefore chooses option 2. The expected utility of attention strategy $(\not\theta, \not s)$ is then:

$$E_{(\not\theta,\not s)} = u_2 = 0.$$

If $E[s] = 0$, the agents expects that $E[u_1] = E[\theta]E[s] = 0 = u_2$ and therefore is indifferent between option 1 and option 2. She randomizes and her expected utility is:

$$E_{(\not\theta,\not s)} = 0.5E[u_1] + 0.5u_2 = 0.$$

If $E[s] > 0$, the agent expects that $E[u_1] = E[\theta]E[s] > 0 = u_2$ and therefore chooses option 1:

$$E_{(\not\theta,\not s)} = E[u_1] = E[\theta]E[s].$$

Attention Strategy $(\theta, (\not s, \not s))$**:** The agent always processes θ, but never processes s. Thus the agent does not know which option is better and decides between options 1 and 2 given the expected value of s. Nevertheless she still has costs for processing θ. If $E[s] < 0$, she expects that $E[u_1] < 0 = u_2$. Therefore, she chooses option 2. The expected utility is :

$$E_{(\theta,(\not s,\not s))} = u_2 - c_\theta = -c_\theta.$$

If $E[s] = 0$, the agent expects that $E[u_1] = 0 = u_2$. She is indifferent between options 1 and 2. She randomizes and her expected utility is:

$$E_{(\theta,(\not s,\not s))} = p_\theta(0.5E[u_1] + 0.5u_2) + (1-p_\theta)(0.5E[u_1] + 0.5u_2) - c_\theta = -c_\theta.$$

If $E[s] > 0$, the agent expects that $E[u_1] > 0 = u_2$. Therefore, she chooses option 1:

$$E_{(\theta,(\not s,\not s))} = p_\theta E[u_1] + (1-p_\theta)E[u_1] - c_\theta = E[\theta]E[s] - c_\theta.$$

Attention Strategy $(\theta, (\not{s}, s))$**:** The agent processes θ and if $\theta = \theta^L$ also processes s. If $s = s^+$, she chooses option 1, and if $s = s^-$, she chooses option 2. If $\theta = \theta^H$, the agent does not process s and thus decides given $E[s]$. As the agent processes s only if $\theta = \theta^H$, she incurs costs for processing s with probability $1 - p_\theta$. But, she always incurs costs for processing θ. If $E[s] < 0$:

$$\begin{aligned} E_{(\theta,(\not{s},s))} &= (1 - p_\theta)\left(p_s u_1 + (1 - p_s)u_2\right) + p_\theta u_2 - c_\theta - (1 - p_\theta)c_s \\ &= (1 - p_\theta)p_s \theta^L s^+ - c_\theta - (1 - p_\theta)c_s. \end{aligned}$$

If $E[s] = 0$:

$$\begin{aligned} E_{(\theta,(\not{s},s))} &= (1 - p_\theta)\left(p_s u_1 + (1 - p_s)u_2\right) + p_\theta(0.5E[u_1] + 0.5u_2) - c_\theta - (1 - p_\theta)c_s \\ &= (1 - p_\theta)p_s \theta^L s^+ - c_\theta - (1 - p_\theta)c_s. \end{aligned}$$

If $E[s] > 0$:

$$\begin{aligned} E_{(\theta,(\not{s},s))} &= (1 - p_\theta)\left(p_s u_1 + (1 - p_s)u_2\right) + p_\theta E[u_1] - c_\theta - (1 - p_\theta)c_s \\ &= (1 - p_\theta)p_s \theta^L s^+ + p_\theta \theta^H E[s] - c_\theta - (1 - p_\theta)c_s. \end{aligned}$$

Attention Strategy $(\theta, (s, \not{s}))$**:** The agent always processes θ, but processes s only if $\theta = \theta^H$. Then, if $s = s^+$, she chooses option 1 ($u_1 = \theta^H s^+ > 0 = u_2$), and if $s = s^-$, she chooses option 2 ($u_1 = \theta^H s^- < 0 = u_2$). If $\theta = \theta^L$, the agent does not process s and decides given her expectations about s. As the agent process s only if $\theta = \theta^H$, she incurs costs for processing s with the probability that $\theta = \theta^H$, i.e., p_θ. But, she always has costs for processing θ. If $E[s] < 0$:

$$E_{(\theta,(s,\not{s}))} = p_\theta\left(p_s u_1 + (1 - p_s)\, u_2\right) + (1 - p_\theta)u_2 - c_\theta - p_\theta c_s = p_\theta p_s \theta^H s^+ - c_\theta - p_\theta c_s.$$

If $E[s] = 0$:

$$\begin{aligned} E_{(\theta,(s,\not{s}))} &= p_\theta\left(p_s u_1 + (1 - p_s)u_2\right) + (1 - p_\theta)(0.5E[u_1] + 0.5u_2) - c_\theta - p_\theta c_s \\ &= p_\theta p_s \theta^H s^+ - c_\theta - p_\theta c_s. \end{aligned}$$

If $E[s] > 0$:

$$\begin{aligned} E_{(\theta,(s,\not{s}))} &= p_\theta\left(p_s u_1 + (1 - p_s)u_2\right) + (1 - p_\theta)E[u_1] - c_\theta - p_\theta c_s \\ &= p_\theta p_s \theta^H s^+ + (1 - p_\theta)\theta^L E[s] - c_\theta - p_\theta c_s. \end{aligned}$$

B Proof of Proposition 1

The proof proceeds by comparing the expected utilities of the attention strategies. As the expected utilities of the attention strategies $(\theta, (\not s, \not s))$ and $(\theta, (\not s, s))$ depend on the expected value of s, we need to distinguish two cases for part (b) and (e) of the proof:[21]

(a) Strategy (θ, s) is strictly inferior to strategy $(s, (\not\theta, \not\theta))$.[22]

As $E_{(s,(\not\theta,\not\theta))} = E[\theta]p_s s^+ - c_s$ and $E_{(\theta,s)} = E[\theta]p_s s^+ - c_s - c_\theta$, $E_{(s,(\not\theta,\not\theta))} > E_{(\theta,s)}$ if $c_\theta > 0$, which is true by assumption.

(b) Strategy $(\theta, (\not s, \not s))$ is strictly inferior to strategy $(\not\theta, \not s)$.

Case $E[s] \leq 0$: As $E_{(\not\theta,\not s)} = 0$ and $E_{(\theta,(\not s,\not s))} = -c_\theta$, $E_{(\not\theta,\not s)} > E_{(\theta,(\not s,\not s))}$ if $c_\theta > 0$, which is true by assumption.

Case $E[s] > 0$: As $E_{(\not\theta,\not s)} = E[\theta]E[s]$ and $E_{(\theta,(\not s,\not s))} = E[\theta]E[s] - c_\theta$, $E_{(\not\theta,\not s)} > E_{(\theta,(\not s,\not s))}$ if $c_\theta > 0$, which is true by assumption.

(c) Strategy $(s, (\theta, \not\theta))$ is never optimal. If $p_s \neq 0$, $E_{(s,(\not\theta,\not\theta))} = E[\theta]p_s s^+ - c_s$ and $E_{(s,(\theta,\not\theta))} = E[\theta]p_s s^+ - c_s - p_s c_\theta$. Thus $E_{(s,(\not\theta,\not\theta))} > E_{(s,(\theta,\not\theta))}$ if $c_\theta > 0$, which is true by assumption. If $p_s = 0 \Leftrightarrow E[s] = s^- < 0$, $E_{(\not\theta,\not s)} = 0$. Then, $E_{(\not\theta,\not s)} > E_{(s,(\theta,\not\theta))}$, if $c_s > 0$, which is true by assumption.

(d) Strategy $(s, (\not\theta, \theta))$ never optimal. If $p_s \neq 1$, $E_{(s,(\not\theta,\not\theta))} = E[\theta]p_s s^+ - c_s$ and $E_{(s,(\not\theta,\theta))} = E[\theta]p_s s^+ - c_s - (1 - p_s)c_\theta$. Thus $E_{(s,(\not\theta,\not\theta))} > E_{(s,(\not\theta,\theta))}$ if $c_\theta > 0$, which is true by assumption. If $p_s = 1 \Leftrightarrow E[s] = s^+ > 0$, $E_{(\not\theta,\not s)} = E[\theta]E[s]$. Then, $E_{(\not\theta,\not s)} > E_{(s,(\not\theta,\theta))}$ if $c_s > 0$, which is true by assumption.

(e) Strategy $(\theta, (\not s, s))$ is never optimal. By contradiction. Attention strategy $(\theta, (\not s, s))$ is optimal if it yields at least as much utility as any other strategy, i.e., $E_{(\theta,(\not s,s))} \geq E_j \; \forall \; j \in A \setminus \{(\theta, (\not s, s))\}$. If $E_{(\theta,(\not s,s))} \geq E_{(\not\theta,\not s)} \wedge E_{(\theta,(\not s,s))} \geq E_{(s,(\not\theta,\not\theta))}$, by part (a) of this proof $E_{(\theta,(\not s,s))} \geq E_{(\theta,s)}$, by part (b) $E_{(\theta,(\not s,s))} \geq E_{(\theta,(\not s,\not s))}$, by (c) $E_{(\theta,(\not s,s))} \geq E_{(s,(\theta,\not\theta))}$, and by part (d) $E_{(\theta,(\not s,s))} \geq E_{(s,(\not\theta,\theta))}$. Thus for strategy $(\theta, (\not s, s))$ to be optimal, it suffices that $E_{(\theta,(\not s,s))} \geq E_{(\not\theta,\not s)} \wedge E_{(\theta,(\not s,s))} \geq E_{(s,(\not\theta,\not\theta))} \wedge E_{(\theta,(\not s,s))} \geq E_{(\theta,(s,\not s))}$.

[21] See Appendix A for an elaborate derivation of the expected utilities.

[22] If a strategy is strictly inferior to another strategy, it cannot be optimal.

Case $E[s] \leq 0$**:**

$$E_{(\theta,(\not s,s))} \geq E_{(\not\theta,\not s)} \qquad \Leftrightarrow \quad c_\theta \leq (1-p_\theta)p_s\theta^L s^+ - (1-p_\theta)c_s. \tag{1a}$$

$$E_{(\theta,(\not s,s))} \geq E_{(s,(\not\theta,\not\theta))} \qquad \Leftrightarrow \quad c_\theta \leq -p_\theta p_s\theta^H s^+ + p_\theta c_s. \tag{1b}$$

$$E_{(\theta,(\not s,s))} \geq E_{(\theta,(s,\not s))} \qquad \Leftrightarrow \begin{cases} c_s \leq \frac{p_s s^+\left((1-p_\theta)\theta^L - p_\theta\theta^H\right)}{1-2p_\theta} & \text{if } p_\theta < 0.5 \\ \theta^H \leq \theta^L & \text{if } p_\theta = 0.5 \\ c_s \geq \frac{p_s s^+\left((1-p_\theta)\theta^L - p_\theta\theta^H\right)}{1-2p_\theta} & \text{if } p_\theta > 0.5. \end{cases} \tag{1c}$$

If conditions (1a)–(1c) are fulfilled, strategy $(\theta,(\not s,s))$ is optimal. However, if $p_\theta = 0.5$, condition (1c) is never fulfilled, because per definition $\theta^H > \theta^L$.

If $p_\theta < 0.5$ and (1c) holds, (1b) cannot hold. If (1c) holds, for (1b)

$$\begin{aligned} c_\theta &\leq -p_\theta p_s\theta^H s^+ + p_\theta c_s \\ &\leq -p_\theta p_s\theta^H s^+ + p_\theta \frac{p_s s^+\left((1-p_\theta)\theta^L - p_\theta\theta^H\right)}{1-2p_\theta} \\ &= p_\theta p_s s^+(1-p_\theta)\frac{\theta^L-\theta^H}{1-2p_\theta} \leq 0. \end{aligned}$$

Therefore if (1c) holds, by (1b) $c_\theta \leq 0$. This is a contradiction. Thus if (1c) holds, condition (1b) is never fulfilled.

If $p_\theta > 0.5$ and (1c) holds, (1a) cannot hold. If (1c) holds, for (1a)

$$\begin{aligned} c_\theta &\leq (1-p_\theta)p_s\theta^L s^+ - (1-p_\theta)c_s \\ &\leq (1-p_\theta)\left(p_s\theta^L s^+ - \frac{p_s s^+\left((1-p_\theta)\theta^L - p_\theta\theta^H\right)}{1-2p_\theta}\right) \\ &= (1-p_\theta)p_s s^+ p_\theta\frac{\theta^H-\theta^L}{1-2p_\theta} \leq 0. \end{aligned}$$

Therefore, if (1c) holds, by (1a) $c_\theta \leq 0$. This is a contradiction. Thus if (1c) holds, condition (1a) is never fulfilled.

Thus, if $E[s] \leq 0$, no case exists such that all three conditions hold at once. Consequently, if $E[s] \leq 0$, strategy $(\theta,(\not s,s))$ is never optimal.

Case $E[s] > 0$:

$$E_{(\theta,(\not{s},s))} \geq E_{(\not{\theta},\not{s})} \quad \Leftrightarrow \quad c_\theta \leq -(1-p_\theta)(1-p_s)\theta^L s^- - (1-p_\theta)c_s \tag{2a}$$

$$E_{(\theta,(\not{s},s))} \geq E_{(s,(\not{\theta},\not{\theta}))} \quad \Leftrightarrow \quad c_\theta \leq p_\theta(1-p_s)\theta^H s^- + p_\theta c_s \tag{2b}$$

$$E_{(\theta,(\not{s},s))} \geq E_{(\theta,(s,\not{s}))} \quad \Leftrightarrow \begin{cases} c_s \leq \frac{(1-p_s)s^-\left(p_\theta\theta^H-(1-p_\theta)\theta^L\right)}{1-2p_\theta} & \text{if } p_\theta < 0.5 \\ \theta^H \leq \theta^L & \text{if } p_\theta = 0.5 \\ c_s \geq \frac{(1-p_s)s^-\left(p_\theta\theta^H-(1-p_\theta)\theta^L\right)}{1-2p_\theta} & \text{if } p_\theta > 0.5 \end{cases} \tag{2c}$$

If conditions (2a)–(2c) are fulfilled, strategy $(\theta, (\not{s}, s))$ is optimal. However, if $p_\theta = 0.5$, condition (2c) is never fulfilled, because per definition $\theta^H > \theta^L$.

If $p_\theta < 0.5$ and (2c) holds, (2b) cannot hold. If (2c) holds, for (2b)

$$\begin{aligned} c_\theta &\leq p_\theta(1-p_s)\theta^H s^- + p_\theta c_s \\ &\leq p_\theta(1-p_s)\theta^H s^- + p_\theta \frac{(1-p_s)s^-\left(p_\theta\theta^H-(1-p_\theta)\theta^L\right)}{1-2p_\theta} \\ &= (1-p_\theta)p_\theta(1-p_s)s^- \frac{\theta^H-\theta^L}{1-2p_\theta} \leq 0. \end{aligned}$$

Therefore, if (2c) holds, by (2b) $c_\theta \leq 0$. This is a contradiction. Thus if (2c) holds, condition (2b) is never fulfilled.

If $p_\theta > 0.5$ and (2c) holds, (2a) cannot hold. If (2c) holds, for (2a)

$$\begin{aligned} c_\theta &\leq -(1-p_\theta)(1-p_s)\theta^L s^- - (1-p_\theta)c_s \\ &\leq -(1-p_\theta)(1-p_s)\theta^L s^- - (1-p_\theta)\frac{(1-p_s)s^-\left(p_\theta\theta^H-(1-p_\theta)\theta^L\right)}{1-2p_\theta} \\ &= (1-p_\theta)p_\theta(1-p_s)s^- \frac{\theta^L-\theta^H}{1-2p_\theta} \leq 0. \end{aligned}$$

Therefore if (2c) holds, by (2a) $c_\theta \leq 0$. This is a contradiction. Thus if (2c) holds, condition (2a) is never fulfilled.

Thus, if $E[s] > 0$, no case exists such that all three conditions hold at once. Consequently, if $E[s] > 0$, strategy $(\theta, (\not{s}, s))$ is never optimal.

Overall, strategy $(\theta, (\not{s}, s))$ is never optimal.

□

C Proof of Proposition 2

The proof proceeds by comparing the expected utilities of the attention strategies. An attention strategy is strictly optimal if it yields more utility than any other strategy. By Proposition 1, it suffices to compare strategies $(\not\theta, \not s)$, $(s, (\not\theta, \not\theta))$, and $(\theta, (s, \not s))$.

Let

$$\underline{c}_\theta = \begin{cases} p_s s^+ (1-p_\theta) p_\theta (\theta^H - \theta^L) & \text{if } E[s] \leq 0 \\ (1-p_s) s^- (1-p_\theta) p_\theta (\theta^L - \theta^H) & \text{if } E[s] > 0. \end{cases} \tag{3}$$

If $p_\theta = 0$ or $p_\theta = 1$, $\underline{c}_\theta = 0$. By assumption $c_\theta > 0$. Therefore, for $p_\theta = 0$ or $p_\theta = 1$, $c_\theta > \underline{c}_\theta$. So for these cases it suffices to show that there exists a $\hat{\kappa}$ such that

i) if $c_s < \hat{\kappa}$, strategy $(s, (\not\theta, \not\theta))$ is strictly optimal,

ii) if $c_s > \hat{\kappa}$, strategy $(\not\theta, \not s)$ is strictly optimal.

Proof of I)

Let $c_\theta \geq \underline{c}_\theta$ be arbitrary. Let

$$\hat{\kappa} = \begin{cases} E[\theta] p_s s^+ & \text{if } E[s] \leq 0 \\ -E[\theta](1-p_s) s^- & \text{if } E[s] > 0 \end{cases}. \tag{4}$$

Assume $c_s < \hat{\kappa}$.
Case 1: Assume $E[s] \leq 0$ such that $\underline{c}_\theta = p_s s^+ (1-p_\theta) p_\theta (\theta^H - \theta^L)$ and $\hat{\kappa} = E[\theta] p_s s^+$. Then

$$\begin{aligned} E_{(s,(\not\theta,\not\theta))} &= E[\theta] p_s s^+ - c_s > E[\theta] p_s s^+ - E[\theta] p_s s^+ = 0 = E_{(\not\theta, \not s)} \\ E_{(s,(\not\theta,\not\theta))} &= E[\theta] p_s s^+ - c_s \\ &= p_\theta p_s \theta^H s^+ + (1-p_\theta) p_s \theta^L s^+ - c_s - c_\theta + c_\theta - p_\theta c_s + p_\theta c_s \\ &= p_\theta p_s \theta^H s^+ - c_\theta - p_\theta c_s + c_\theta - (1-p_\theta) c_s + p_s s^+ (1-p_\theta) \theta^L \\ &\geq E_{(\theta,(s,\not s))} + p_s s^+ (1-p_\theta) p_\theta (\theta^H - \theta^L) - (1-p_\theta) c_s + p_s s^+ (1-p_\theta) \theta^L \\ &> E_{(\theta,(s,\not s))} + p_s s^+ (1-p_\theta) p_\theta (\theta^H - \theta^L) - (1-p_\theta) E[\theta] p_s s^+ + p_s s^+ (1-p_\theta) \theta^L \\ &= E_{(\theta,(s,\not s))} \end{aligned}$$

Case 2: Assume $E[s] > 0$ such that $\underline{c}_\theta = (1-p_s)s^-(1-p_\theta)p_\theta(\theta^L - \theta^H)$ and $\hat{\kappa} = -E[\theta](1-p_s)s^-$. Then

$$\begin{aligned}
E_{(s,(\not\theta,\not\theta))} &= E[\theta]p_s s^+ - c_s > E[\theta]p_s s^+ + E[\theta](1-p_s)s^- = E[\theta]E[s] = E_{(\not\theta,\not s)} \\
E_{(s,(\not\theta,\not\theta))} &= E[\theta]p_s s^+ - c_s = E[\theta]p_s s^+ - c_s + p_\theta c_s - p_\theta c_s + c_\theta - c_\theta \\
&= p_\theta p_s \theta^H s^+ + (1-p_\theta)p_s\theta^L s^+ - p_\theta c_s - c_\theta - (1-p_\theta)c_s + c_\theta \\
&> p_\theta p_s \theta^H s^+ + (1-p_\theta)p_s\theta^L s^+ - p_\theta c_s - c_\theta + (1-p_\theta)E[\theta](1-p_s)s^- + c_\theta \\
&\geq p_\theta p_s \theta^H s^+ + (1-p_\theta)p_s\theta^L s^+ - p_\theta c_s - c_\theta + (1-p_\theta)E[\theta](1-p_s)s^- \\
&\quad + (1-p_s)s^-(1-p_\theta)p_\theta(\theta^L - \theta^H) \\
&= p_\theta p_s \theta^H s^+ + (1-p_\theta)\theta^L E[s] - p_\theta c_s - c_\theta = E_{(\theta,(s,\not s))}
\end{aligned}$$

Thus if $c_s < \hat{\kappa}$, strategy $(s,(\not\theta,\not\theta))$ is strictly optimal.

Assume $c_s > \hat{\kappa}$.

Case 1: Assume $E[s] \leq 0$ such that $\underline{c}_\theta = p_s s^+(1-p_\theta)p_\theta(\theta^H - \theta^L)$ and $\hat{\kappa} = E[\theta]p_s s^+$. Then

$$\begin{aligned}
E_{(\not\theta,\not s)} &= 0 = E[\theta]p_s s^+ - E[\theta]p_s s^+ > E[\theta]p_s s^+ - c_s = E_{(s,(\not\theta,\not\theta))} \\
E_{(\not\theta,\not s)} &= 0 = p_\theta p_s \theta^H s^+ - p_\theta p_s \theta^H s^+ + p_\theta p_\theta p_s \theta^H s^+ - p_\theta p_\theta p_s \theta^H s^+ + (1-p_\theta)p_\theta p_s \theta^L s^+ \\
&\quad - (1-p_\theta)p_\theta p_s \theta^L s^+ \\
&= p_\theta p_s \theta^H s^+ - p_s s^+(1-p_\theta)p_\theta(\theta^H - \theta^L) - p_\theta E[\theta]p_s s^+ \\
&\geq p_\theta p_s \theta^H s^+ - c_\theta - p_\theta E[\theta]p_s s^+ \\
&> p_\theta p_s \theta^H s^+ - c_\theta - p_\theta c_s = E_{(\theta,(s,\not s))}
\end{aligned}$$

Case 2: Assume $E[s] > 0$ such that $\underline{c}_\theta = (1-p_s)s^-(1-p_\theta)p_\theta(\theta^L - \theta^H)$ and $\hat{\kappa} = -E[\theta](1-p_s)s^-$. Then

$$\begin{aligned}
E_{(\not\theta,\not s)} &= E[\theta]E[s] = E[\theta]p_s s^+ + E[\theta](1-p_s)s^- > E[\theta]p_s s^+ - c_s = E_{(s,(\not\theta,\not\theta))} \\
E_{(\not\theta,\not s)} &= E[\theta]E[s] = p_\theta p_s \theta^H s^+ + (1-p_\theta)\theta^L E[s] + p_\theta(1-p_s)\theta^H s^- \\
&= p_\theta p_s \theta^H s^+ + (1-p_\theta)\theta^L E[s] + p_\theta(1-p_s)s^- \left((1-p_\theta)\theta^H + p_\theta\theta^H + (1-p_\theta)\theta^L - (1-p_\theta)\theta^L\right) \\
&= p_\theta p_s \theta^H s^+ + (1-p_\theta)\theta^L E[s] - (1-p_s)s^-(1-p_\theta)p_\theta(\theta^L - \theta^H) + p_\theta E[\theta](1-p_s)s^- \\
&> p_\theta p_s \theta^H s^+ + (1-p_\theta)\theta^L E[s] - c_\theta - p_\theta c_s = E_{(\theta,(s,\not s))}
\end{aligned}$$

Thus if $c_s > \hat{\kappa}$, strategy $(\not\theta,\not s)$ is strictly optimal.

Consequently, as $c_\theta \geq \underline{c}_\theta$ was arbitrary, for all $c_\theta \geq \underline{c}_\theta$, there exists a $\hat{\kappa}$ such that

i) if $c_s < \hat{\kappa}$, strategy $(s,(\not\theta,\not\theta))$ is strictly optimal,

ii) if $c_s > \hat{\kappa}$, strategy $(\not\theta,\not s)$ is strictly optimal,

Proof of II)

Let $c_\theta < \underline{c}_\theta$ be arbitrary. Let

$$\underline{\kappa} = \begin{cases} p_s\theta^L s^+ + \frac{c_\theta}{1-p_\theta} & \text{if } E[s] \leq 0 \\ -(1-p_s)\theta^L s^- + \frac{c_\theta}{1-p_\theta} & \text{if } E[s] > 0, \end{cases} \tag{5}$$

$$\overline{\kappa} = \begin{cases} p_s\theta^H s^+ - \frac{c_\theta}{p_\theta} & \text{if } E[s] \leq 0 \\ -(1-p_s)\theta^H s^- - \frac{c_\theta}{p_\theta} & \text{if } E[s] > 0 \end{cases}. \tag{6}$$

Assume $c_s < \underline{\kappa}$.
Case 1: Assume $E[s] \leq 0$ such that $\underline{c}_\theta = p_s s^+(1-p_\theta)p_\theta(\theta^H - \theta^L)$ and $\underline{\kappa} = p_s\theta^L s^+ + \frac{c_\theta}{1-p_\theta}$. Then

$$\begin{aligned} E_{(s,(\not\theta,\not\theta))} &= E[\theta]p_s s^+ - c_s > E[\theta]p_s s^+ - p_s\theta^L s^+ - \frac{c_\theta}{1-p_\theta} \\ &> p_s s^+ p_\theta(\theta^H - \theta^L) - p_s s^+ p_\theta(\theta^H - \theta^L) = 0 = E_{(\not\theta,\not s)} \\ E_{(s,(\not\theta,\not\theta))} &= E[\theta]p_s s^+ - c_s = E[\theta]p_s s^+ - c_s + p_\theta c_s - p_\theta c_s \\ &> E[\theta]p_s s^+ - p_\theta c_s - (1-p_\theta)\left(p_s\theta^L s^+ + \frac{c_\theta}{1-p_\theta}\right) \\ &= p_\theta p_s\theta^H s^+ + (1-p_\theta)p_s\theta^L s^+ - p_\theta c_s - (1-p_\theta)p_s\theta^L s^+ - c_\theta \\ &= p_\theta p_s\theta^H s^+ - c_\theta - p_\theta c_s = E_{(\theta,(s,\not s))} \end{aligned}$$

Case 2: Assume $E[s] > 0$ such that $\underline{c}_\theta = (1-p_s)s^-(1-p_\theta)p_\theta(\theta^L - \theta^H)$ and $\underline{\kappa} = -(1-p_s)\theta^L s^- + \frac{c_\theta}{1-p_\theta}$. Then

$$\begin{aligned} E_{(s,(\not\theta,\not\theta))} &= E[\theta]p_s s^+ - c_s > E[\theta]p_s s^+ + (1-p_s)\theta^L s^- - \frac{c_\theta}{1-p_\theta} \\ &> E[\theta]p_s s^+ + (1-p_s)\theta^L s^- - (1-p_s)s^- p_\theta(\theta^L - \theta^H) = E[\theta]E[s] = E_{(\not\theta,\not s)} \\ E_{(s,(\not\theta,\not\theta))} &= E[\theta]p_s s^+ - c_s = E[\theta]p_s s^+ - c_s + p_\theta c_s - p_\theta c_s \\ &> E[\theta]p_s s^+ - p_\theta c_s - (1-p_\theta)\left(-(1-p_s)\theta^L s^- + \frac{c_\theta}{1-p_\theta}\right) \\ &= p_\theta p_s\theta^H s^+ + (1-p_\theta)\theta^L E[s] - c_\theta - p_\theta c_s = E_{(\theta,(s,\not s))} \end{aligned}$$

Thus if $c_s < \underline{\kappa}$, strategy $(s,(\not\theta,\not\theta))$ is strictly optimal.

Assume $\underline{\kappa} < c_s < \overline{\kappa}$.

Case 1: Assume $E[s] \leq 0$ such that $\underline{c_\theta} = p_s s^+(1-p_\theta)p_\theta(\theta^H - \theta^L)$, $\underline{\kappa} = p_s\theta^L s^+ + \frac{c_\theta}{1-p_\theta}$, and $\overline{\kappa} = p_s\theta^H s^+ - \frac{c_\theta}{p_\theta}$. Then

$$\begin{aligned}
E_{(\theta,(s,\not{s}))} &= p_\theta p_s\theta^H s^+ - c_\theta - p_\theta c_s \\
&> p_\theta p_s\theta^H s^+ - c_\theta - p_\theta\left(p_s\theta^H s^+ - \frac{c_\theta}{p_\theta}\right) = 0 = E_{(\not{\theta},\not{s})} \\
E_{(\theta,(s,\not{s}))} &= p_\theta p_s\theta^H s^+ - c_\theta - p_\theta c_s = p_\theta p_s\theta^H s^+ - c_\theta - p_\theta c_s + c_s - c_s \\
&> p_\theta p_s\theta^H s^+ - c_\theta - c_s + (1-p_\theta)\left(p_s\theta^L s^+ + \frac{c_\theta}{1-p_\theta}\right) \\
&= E[\theta]p_s s^+ - c_s = E_{(s,(\not{\theta},\not{\theta}))}
\end{aligned}$$

Case 2: Assume $E[s] > 0$ such that $\underline{c_\theta} = (1-p_s)s^-(1-p_\theta)p_\theta(\theta^L - \theta^H)$, $\underline{\kappa} = -(1-p_s)\theta^L s^- + \frac{c_\theta}{1-p_\theta}$, and $\overline{\kappa} = -(1-p_s)\theta^H s^- - \frac{c_\theta}{p_\theta}$. Then

$$\begin{aligned}
E_{(\theta,(s,\not{s}))} &= p_\theta p_s\theta^H s^+ + (1-p_\theta)\theta^L E[s] - c_\theta - p_\theta c_s \\
&> p_\theta p_s\theta^H s^+ + (1-p_\theta)\theta^L E[s] - c_\theta - p_\theta\left(-(1-p_s)\theta^H s^- - \frac{c_\theta}{p_\theta}\right) \\
&= E[\theta]E[s] = E_{(\not{\theta},\not{s})} \\
E_{(\theta,(s,\not{s}))} &= p_\theta p_s\theta^H s^+ + (1-p_\theta)\theta^L E[s] - c_\theta - p_\theta c_s \\
&= E[\theta]p_s s^+ + (1-p_\theta)\theta^L(1-p_s)s^- - c_\theta - p_\theta c_s - c_s + c_s \\
&> E[\theta]p_s s^+ + (1-p_\theta)\theta^L(1-p_s)s^- - c_s - c_\theta \\
&\quad + (1-p_\theta)\left(-(1-p_s)\theta^L s^- + \frac{c_\theta}{1-p_\theta}\right) \\
&= E[\theta]p_s s^+ - c_s = E_{(s,(\not{\theta},\not{\theta}))}
\end{aligned}$$

Thus if $\underline{\kappa} < c_s < \overline{\kappa}$, strategy $(\theta,(s,\not{s}))$ is strictly optimal.

Assume $c_s > \overline{\kappa}$.

Case 1: Assume $E[s] \leq 0$ such that $\underline{c}_\theta = p_s s^+(1-p_\theta)p_\theta(\theta^H - \theta^L)$ and $\overline{\kappa} = p_s\theta^H s^+ - \frac{c_\theta}{p_\theta}$. Then

$$\begin{aligned}
E_{(\not\theta,\not s)} &= 0 = p_s s^+(1-p_\theta)(\theta^L - \theta^H) + p_s s^+(1-p_\theta)(\theta^H - \theta^L) \\
&> E[\theta]p_s s^+ - p_s\theta^H s^+ + \frac{c_\theta}{p_\theta} > E[\theta]p_s s^+ - c_s = E_{(s,(\not\theta,\not\theta))} \\
E_{(\not\theta,\not s)} &= 0 = p_\theta p_s\theta^H s^+ - p_\theta p_s\theta^H s^+ - c_\theta + c_\theta \\
&= p_\theta p_s\theta^H s^+ - c_\theta - p_\theta\left(p_s\theta^H s^+ - \frac{c_\theta}{p_\theta}\right) \\
&> p_\theta p_s\theta^H s^+ - c_\theta - p_\theta c_s = E_{(\theta,(s,\not s))}
\end{aligned}$$

Case 2: Assume $E[s] > 0$ such that $\underline{c}_\theta = (1-p_s)s^-(1-p_\theta)p_\theta(\theta^L - \theta^H)$ and $\overline{\kappa} = -(1-p_s)\theta^H s^- - \frac{c_\theta}{p_\theta}$. Then

$$\begin{aligned}
E_{(\not\theta,\not s)} &= E[\theta]E[s] = E[\theta]p_s s^+ + (1-p_s)\theta^H s^- + (1-p_s)s^-(1-p_\theta)(\theta^L - \theta^H) \\
&> E[\theta]p_s s^+ + (1-p_s)\theta^H s^- + \frac{c_\theta}{p_\theta} > E[\theta]p_s s^+ - c_s = E_{(s,(\not\theta,\not\theta))} \\
E_{(\not\theta,\not s)} &= E[\theta]E[s] = p_\theta\theta^H E[s] + (1-p_\theta)\theta^L E[s] - c_\theta + c_\theta \\
&= p_\theta p_s\theta^H s^+ + (1-p_\theta)\theta^L E[s] - c_\theta - p_\theta\left(-\theta^H(1-p_s)s^- - \frac{c_\theta}{p_\theta}\right) \\
&> p_\theta p_s\theta^H s^+ + (1-p_\theta)\theta^L E[s] - c_\theta - p_\theta c_s = E_{(\theta,(s,\not s))}
\end{aligned}$$

Thus if $c_s > \overline{\kappa}$, strategy $(\not\theta, \not s)$ is strictly optimal.

Consequently, as $c_\theta < \underline{c}_\theta$ was arbitrary, for all $c_\theta < \underline{c}_\theta$, there exists a $\underline{\kappa}$ and a $\overline{\kappa}$ such that

i) if $c_s < \underline{\kappa}$, strategy $(s, (\not\theta, \not\theta))$ is strictly optimal,

ii) if $\underline{\kappa} < c_s < \overline{\kappa}$, strategy $(\theta, (s, \not s))$ is strictly optimal,

iii) if $c_s > \overline{\kappa}$, strategy $(\not\theta, \not s)$ is strictly optimal.

Overall thus there exists a $\underline{c}_\theta$ such that Proposition 2 holds.

□

D Proof of Lemma 1

The proof proceeds in four parts, following the distinction in Lemma 1:

i) As both lines of (3) increase with θ^H, $\underline{c}_\theta$ increases with θ^H. As the first line of (3) (case $E[s] \leq 0$) increases with s^+ and p_s, $\underline{c}_\theta$ increases with s^+ and p_s if $E[s] \leq 0$. As both lines of (3) decrease with θ^L, $\underline{c}_\theta$ decreases with θ^L. As the second line of (3) (case $E[s] > 0$) with s^- and p_s, $\underline{c}_\theta$ decreases with s^- and p_s if $E[s] > 0$. As both lines of (3) increase (decrease) with p_θ if $p_\theta < 0.5$ ($p_\theta > 0.5$), $\underline{c}_\theta$ increases (decreases) with p_θ if $p_\theta < 0.5$ ($p_\theta > 0.5$).

ii) As both lines of (4) increase with θ^H, θ^L, and p_θ, $\hat{\kappa}$ increases with θ^H, θ^L, and p_θ. As the first line of (4) (case $E[s] \leq 0$) increases with s^+ and p_s, $\hat{\kappa}$ increases with s^+ and p_s if $E[s] \leq 0$. As the second line of (4) (case $E[s] > 0$) decreases with s^- and p_s, $\hat{\kappa}$ decreases with s^- and p_s if $E[s] > 0$.

iii) As both lines of (5) increase with θ^L, p_θ, and c_θ, $\underline{\kappa}$ increases with θ^L, p_θ, and c_θ. As the first line of (5) (case $E[s] \leq 0$) increases with s^+ and p_s, $\underline{\kappa}$ increases with s^+ and p_s if $E[s] \leq 0$. As the second line of (5) (case $E[s] > 0$) decreases with s^- and p_s, $\underline{\kappa}$ decreases with s^- and p_s if $E[s] > 0$.

iv) As both lines of (6) increase with θ^H and p_θ, $\overline{\kappa}$ increases with θ^H and p_θ. As the first line of (6) (case $E[s] \leq 0$) increases with s^+ and p_s, $\overline{\kappa}$ increases with s^+ and p_s if $E[s] \leq 0$. As both lines of (6) decrease with c_θ, $\overline{\kappa}$ decreases with c_θ. As the second line of (6) (case $E[s] > 0$) decreases with s^- and p_s, $\overline{\kappa}$ decreases with s^- and p_s if $E[s] > 0$.

□

E Proof of Remark 1

For the proof two cases have to be differentiated: $E[s] < 0$ and $E[s] > 0$.

Assume $E[s] < 0$ and $\rho > k$. Then if the consumers do not process s_i, they choose option 2—not to buy. Thus with strategy $(\not\theta, \not s)$ and with strategy $(\theta, (s, \not s))$ when $\theta = \theta^L$ demand is $x_i^D = 0$. If consumers process s_i and $s_i = s^+$, consumers choose option 1—to buy. Thus if $s_i = s^+$, with strategies $(s, (\not\theta, \not\theta))$ and $(\theta, (s, \not s))$ when $\theta = \theta^H$ demand is $x_i^D = 1 \cdot r_i$. If $s_i = s^-$ the demand respectively is $x_i^D = 0$. Overall, the profit is

$$\Pi_i = (\rho - K(s_i))x_i^D = \begin{cases} 0 & \text{if } (\not\theta, \not s) \text{ and } s_i = s^+ \\ 0 & \text{if } (\not\theta, \not s) \text{ and } s_i = s^- \\ (\rho - k)r_i & \text{if } (s, (\not\theta, \not\theta)) \text{ and } s_i = s^+ \\ 0 & \text{if } (s, (\not\theta, \not\theta)) \text{ and } s_i = s^- \\ (\rho - k)r_i p_\theta & \text{if } (\theta, (s, \not s)) \text{ and } s_i = s^+ \\ 0 & \text{if } (\theta, (s, \not s)) \text{ and } s_i = s^-. \end{cases} \tag{7}$$

The profit is maximal if consumers use $(s, (\not\theta, \not\theta))$ and $s_i = s^+$.[23] Consumers choose strategy $(s, (\not\theta, \not\theta))$, if $\underline{c}_\theta > c_\theta$ and $c_s < \underline{\kappa}$ or if $\underline{c}_\theta \leq c_\theta$ and $c_s < \hat{\kappa}$ (see Proposition 2). So if $E[s] < 0$ and $\rho > k$, firms have an incentive to produce high quality and highlight this by setting the costs of processing quality such that consumers process quality.

Assume $E[s] > 0$. Then if consumers process $s_i = s^+$ or do not process s_i, they buy the good of firm i. Thus with strategy $(\not\theta, \not s)$ and if $s = s^+$ with strategies $(s, (\not\theta, \not\theta))$ and $(\theta, (s, \not s))$, demand is $x_i^D = 1 \cdot r_i$. If $s = s^-$ demand is zero if quality is processed. Thus with $(s, (\not\theta, \not\theta))$ demand is $x_i^D = 0$ and with $(\theta, (s, \not s))$ demand is $x_i^D = r_i(1 - p_\theta)$.

The profit is thus

$$\Pi_i = (\rho - K(s_i))x_i^D = \begin{cases} (\rho - k)r_i & \text{if } (\not\theta, \not s) \text{ and } s_i = s^+ \\ \rho r_i & \text{if } (\not\theta, \not s) \text{ and } s_i = s^- \\ (\rho - k)r_i & \text{if } (s, (\not\theta, \not\theta)) \text{ and } s_i = s^+ \\ 0 & \text{if } (s, (\not\theta, \not\theta)) \text{ and } s_i = s^- \\ (\rho - k)r_i & \text{if } (\theta, (s, \not s)) \text{ and } s_i = s^+ \\ \rho r_i(1 - p_\theta) & \text{if } (\theta, (s, \not s)) \text{ and } s_i = s^-. \end{cases} \tag{8}$$

[23]Exception: $p_\theta = 1$, but if $p_\theta = 1$, $(\theta, (s, \not s))$ is never optimal and so not inducable by the firm.

If $E[s] > 0$ the profit is maximal if $s_i = s^-$ and consumers use strategy $(\not\theta, \not s)$.[24] Consumers use strategy $(\not\theta, \not s)$ if $\underline{c}_\theta > c_\theta$ and $c_s > \overline{\kappa}$ or $\underline{c}_\theta \leq c_\theta$ and $c_s > \hat{\kappa}$ (see Proposition 2). So if $E[s] > 0$ firms have an incentive to produce low quality and shroud this by setting the costs for processing quality $c_s > \overline{\kappa}$, such that consumers do not process quality.

□

F Proof of Proposition 3

Assume $k > \rho$, then whenever $s_i = s^+$, $\Pi_i = (\rho - k)x_i^D \leq 0$, and whenever $s = s^-$, $\Pi_i = \rho x_i^D \geq 0$. So no firm has an incentive to deviate from producing low quality. If all firms produce low quality, firms make the same profit independent of the attention strategy of the consumers, because consumers expect low quality and choose option 2. Consumers utility is 0 and firms' profit is 0.

Assume $k < \rho < 2k$. If $E[s] = 0$, a fraction p_s of the firms produce high quality and highlight quality, and a fraction $1 - p_s$ of the firms produce low quality and shroud quality. The profit is given as follows: Consumers randomize when they do not process s_i, buy if they process $s_i = s^+$ and do not buy if they process $s_i = s^-$. Thus overall, if $E[s] = 0$, profits are

$$\Pi_i = \begin{cases} (\rho - k)0.5r_i & \text{if } (\not\theta, \not s) \text{ and } s_i = s^+ \\ \rho 0.5 r_i & \text{if } (\not\theta, \not s) \text{ and } s_i = s^- \\ (\rho - k)r_i & \text{if } (s, (\not\theta, \not\theta)) \text{ and } s_i = s^+ \\ 0 & \text{if } (s, (\not\theta, \not\theta)) \text{ and } s_i = s^- \\ (\rho - k)r_i(p_\theta + 0.5(1 - p_\theta)) & \text{if } (\theta, (s, \not s)) \text{ and } s_i = s^+ \\ \rho r_i 0.5(1 - p_\theta) & \text{if } (\theta, (s, \not s)) \text{ and } s_i = s^- \end{cases}$$

Firms using strategy $s = s^+$ and highlight this receive a profit of $(\rho - k)r_i$. These firms have no incentive to deviate to shrouding, because the profit they can make by shrouding is either $(\rho - k)0.5r_i \leq (\rho - k)r_i$ or if they only partially shroud $(\rho - k)r_i(p_\theta + 0.5(1 - p_\theta)) \leq (\rho - k)r_i$. Firms also have no incentive to deviate to producing low quality, because then, they change the distribution of quality and $E[s] < 0$ and by (7), the maximal profit they can receive is $0 < (\rho - k)r_i$.

Firms producing low quality and shrouding this receive a profit of $\rho 0.5 r_i$. These firms have no incentive to deviate. By highlighting their quality, the profit they receive is $0 < \rho 0.5 r_i$ or if they only partially highlight the profit they receive is $\rho r_i 0.5(1 - p_\theta) \leq \rho 0.5 r_i$. By deviating

[24] Exception: $p_\theta = 0$, but if $p_\theta = 0$, $(\theta, (s, \not s))$ is never optimal and so not triggered by the firm.

to high quality they change the distribution of quality and $E[s] > 0$ and by (8) the maximal profit the firm can make is $(\rho - k)r_i < \rho 0.5 r_i$.

So neither has an incentive to deviate.

Assume $\rho > 2k$. If a fraction $p'_s = min\{p_s|E[s] > 0\}$ produce high quality and highlight quality, and a fraction $1 - p'_s$ produce low quality and shroud quality, $E[s] > 0$ and no firm has an incentive to deviate. First, firms producing low quality and shrouding receive a profit of ρr_i. Low quality and shrouding maximizes profits if $E[s] > 0$. Therefore, firms producing low quality and shrouding have no incentive to deviate. Second, the fraction of firms producing high quality and highlighting quality receive a profit of $(\rho - k)r_i$. If they deviate to shrouding, they receive the same payoff. So they have no incentive to deviate to shrouding. If they deviate to producing low quality, they change the distribution of quality and $E[s] = 0$, so that the maximum profit they can receive is $\rho 0.5 r_i$. However, because $\rho > 2k$, $\rho 0.5 r_i < (\rho - k)r_i$. So they have no incentive to deviate.

So neither has an incentive to deviate.

□

BERG Working Paper Series (most recent publications)

82 Björn-Christopher **Witte**, Removing systematic patterns in returns in a financial market model by artificially intelligent traders, October 2011

83 Reiner **Franke** and Frank **Westerhoff**, Why a Simple Herding Model May Generate the Stylized Facts of Daily Returns: Explanation and Estimation, December 2011

84 Frank **Westerhoff**, Interactions between the real economy and the stock market, December 2011

85 Christoph **Wunder** and Guido **Heineck**, Working time preferences, hours mismatch and well-being of couples: Are there spillovers?, October 2012

86 Manfred **Antoni** and Guido **Heineck**, Do literacy and numeracy pay off? On the relationship between basic skills and earnings, October 2012

87 János **Seregi**, Zsuzsanna **Lelovics** and László **Balogh**, The social welfare function of forests in the light of the theory of public goods, October 2012

88 Frank **Westerhoff** and Reiner **Franke**, Agent-based models for economic policy design: two illustrative examples, November 2012

89 Fabio **Tramontana**, Frank **Westerhoff** and Laura **Gardini**, The bull and bear market model of Huang and Day: Some extensions and new results, November 2012

90 Noemi **Schmitt** and Frank **Westerhoff,** Speculative behavior and the dynamics of interacting stock markets, November 2013

91 Jan **Tuinstra**, Michael **Wegener** and Frank **Westerhoff,** Positive welfare effects of trade barriers in a dynamic equilibrium model, November 2013

92 Philipp **Mundt**, Mishael **Milakovic** and Simone **Alfarano,** Gibrat’s Law Redux: Think Profitability Instead of Growth, January 2014

93 Guido **Heineck**, Love Thy Neighbor – Religion and Prosocial Behavior, October 2014

94 Johanna Sophie **Quis**, Does higher learning intensity affect student well-being? Evidence from the National Educational Panel Study, January 2015

95 Stefanie P. **Herber**, The Role of Information in the Application for Merit-Based Scholarships: Evidence from a Randomized Field Experiment, January 2015

96 Noemi **Schmitt** and Frank **Westerhoff**, Managing rational routes to randomness, January 2015

97 Dietmar **Meyer** and Adela **Shera**, Remittances’ Impact on the Labor Supply and on the Deficit of Current Account, February 2015

98 Abdylmenaf **Bexheti** and Besime **Mustafi**, Impact of Public Funding of Education on Economic Growth in Macedonia, February 2015

99 Roberto **Dieci** and Frank **Westerhoff**, Heterogeneous expectations, boom-bust housing cycles, and supply conditions: a nonlinear dynamics approach, April 2015

100 Stefanie P. **Herber**, Johanna Sophie **Quis**, and Guido **Heineck**, Does the Transition into Daylight Saving Time Affect Students' Performance?, May 2015

101 Mafaïzath A. **Fatoke-Dato**, Impact of an educational demand-and-supply policy on girls' education in West Africa: Heterogeneity in income, school environment and ethnicity, June 2015

102 Mafaïzath A. **Fatoke-Dato**, Impact of income shock on children's schooling and labor in a West African country, June 2015

103 Noemi **Schmitt**, Jan **Tuinstra** and Frank **Westerhoff**, Side effects of nonlinear profit taxes in an evolutionary market entry model: abrupt changes, coexisting attractors and hysteresis problems, August 2015.

104 Noemi **Schmitt** and Frank **Westerhoff**, Evolutionary competition and profit taxes: market stability versus tax burden, August 2015.

105 Lena **Dräger** and Christian R. **Proaño**, Cross-Border Banking and Business Cycles in Asymmetric Currency Unions, November 2015.

106 Christian R. **Proaño** and Benjamin **Lojak**, Debt Stabilization and Macroeconomic Volatility in Monetary Unions under Heterogeneous Sovereign Risk Perceptions, November 2015.

107 Noemi **Schmitt** and Frank **Westerhoff**, Herding behavior and volatility clustering in financial markets, February 2016

108 Jutta **Viinikainen**, Guido **Heineck**, Petri **Böckerman**, Mirka **Hintsanen**, Olli **Raitakari** and Jaakko **Pehkonen**, Born Entrepreneur? Adolescents' Personality Characteristics and Self-Employment in Adulthood, March 2016

109 Stefanie P. **Herber** and Michael **Kalinowski**, Non-take-up of Student Financial Aid: A Microsimulation for Germany, April 2016

110 Silke **Anger** and Daniel D. **Schnitzlein**, Cognitive Skills, Non-Cognitive Skills, and Family Background: Evidence from Sibling Correlations, April 2016

111 Noemi **Schmitt** and Frank **Westerhoff**, Heterogeneity, spontaneous coordination and extreme events within large-scale and small-scale agent-based financial market models, June 2016

112 Benjamin **Lojak**, Sentiment-Driven Investment, Non-Linear Corporate Debt Dynamics and Co-Existing Business Cycle Regimes, July 2016

113 Julio **González-Díaz**, Florian **Herold** and Diego **Domínguez**, Strategic Sequential Voting, July 2016

114 Stefanie Yvonne **Schmitt**, Rational Allocation of Attention in Decision-Making, July 2016