

Smith, John

**Working Paper**

## Imperfect memory and the preference for increasing payments

Working Paper, No. 2008-05

**Provided in Cooperation with:**

Department of Economics, Rutgers University

*Suggested Citation:* Smith, John (2008) : Imperfect memory and the preference for increasing payments, Working Paper, No. 2008-05, Rutgers University, Department of Economics, New Brunswick, NJ

This Version is available at:

<https://hdl.handle.net/10419/59490>

**Standard-Nutzungsbedingungen:**

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

**Terms of use:**

*Documents in EconStor may be saved and copied for your personal and scholarly purposes.*

*You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.*

*If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.*

# Imperfect Memory and the Preference for Increasing Payments\*

John Smith<sup>†</sup>  
Rutgers University-Camden

July 31, 2008

## Abstract

In this paper we show how imperfect memory can imply a preference for increasing payments. We model an agent making a decision regarding effort in two periods where the cost of effort is imperfectly known. Before making the first decision, the agent receives a signal related to the cost of effort, which is subsequently forgotten. Before the second decision, the agent makes an inference regarding the content of this signal based on the publicly available information: the action taken and the wage paid. A preference for increasing payments naturally emerges from our model. With the auxiliary assumption that obtaining wage income requires an unknown cost of effort and obtaining rental income requires a known, zero cost of effort, our results provide an explanation for the experimental findings of Loewenstein and Sicherman (1991). These authors find evidence of a stronger preference for increasing "income from wages" rather than "income from rent." Additionally, our model makes the novel prediction that this preference for increasing payments will only occur when the contracts are neither very likely nor very unlikely to cover the cost of effort.

---

\*This paper has benefited from discussions with Roland Benabou, Faruk Gul, Jo Hertel, Marcelo Pinheiro, participants of the Midwest Economics Association Meetings in Minneapolis, the SABE Conference in New York and the Whitebox Advisors Graduate Student Conference at Yale.

<sup>†</sup>Email: [smithj@camden.rutgers.edu](mailto:smithj@camden.rutgers.edu); Website: <http://crab.rutgers.edu/~smithj>; Phone: (856) 225-6319.

# 1 Introduction

In this paper, we model an agent who is to make a decision regarding effort in two periods. Before the first decision, the agent receives a private signal regarding the cost of effort. In the second period, the agent forgets this signal but makes an inference regarding its content from the publicly available facts which are recalled: the action taken and the payment for the action. It is advantageous for the agent to consider the influence of the first period choice of action on second period beliefs. As a result the agent might exhibit a preference for increasing payments.

This implication of our model relates to the choice experiments found in Loewenstein and Sicherman (1991). These authors find that people have a stronger preference for an increasing sequence of payments when described as "income for wages" rather than "income from rent." The authors attribute this result to a feeling of mastery derived from increasing wage payments but not derived from increasing payments of rent. By contrast, we explain this result through a different interpretation of difference between wage payments and rent payments: wages require repeated effort and rent does not. We contend that this finding can be explained by work requiring the repeated expenditure of an unknown, imperfectly recalled cost of effort whereas rent requires a cost of effort known to be zero.

A key aspect of our model is the inference of the forgotten signal. In order to illustrate this inference in a setting without strategic issues, consider the following simple example.

**Example 1** *Suppose an agent is to make a single decision regarding effort, the cost of which is imperfectly known. The cost of effort can either be low or high:  $c \in \{c_L, c_H\}$ , where  $c_L = 0$ ,  $c_H = 1$  and each occurs with equal probability. Before deciding on an action, the agent receives a signal ( $s \in \{s_H, s_L\}$ ) which imperfectly reveals the true cost of effort:*

$$\begin{aligned} P(s_H|c_H) &= P(s_L|c_L) = q \\ P(s_L|c_H) &= P(s_H|c_L) = 1 - q \end{aligned}$$

where  $q \in (0.5, 1)$ . The agent then decides to take the action or not ( $a \in \{1, 0\} = A$ ) where 1 indicates high effort and 0 indicates low effort. In the second period, the agent can only recall the publicly available information:  $a$  and  $w$  but not  $s$ .<sup>1</sup> From the information which is recalled, the agent makes an inference regarding the true cost of effort. We denote the posterior belief of a high cost of effort as  $\bar{\beta}(a, w(a))$ . Consider two contracts,  $w$  and  $w'$ . The contract  $w$  pays  $z$  for high effort ( $w(1) = z$ ) and 0 for low effort ( $w(0) = 0$ ) Contract  $w'$  is

---

<sup>1</sup>Perhaps intuition for this assumption can be best obtained by asking the following three questions to a non-economist friend: "At work today, did you complete the tasks required of you?", "How much were you paid?" and "What was your cost of effort?" The first two questions have rather straightforward answers. The last question will likely provoke a look of bewilderment, because there is not a straightforward answer. If the tasks were completed, we economists infer that the cost of effort was less than the utility from payment, however articulating the precise amount is a much more difficult matter.

such that  $w'(1) = z'$  and  $w'(0) = 0$  where  $z > z'$ . Further suppose that

$$z > q > z' > 1 - q.$$

Under contract  $w$ , the agent will select  $a = 1$  for any signal, however under contract  $w'$  the agent will select  $a = 1$  if  $s = s_L$  and  $a = 0$  if  $s = s_H$ . It follows that  $\bar{\beta}(1, z) = 0.5 > \bar{\beta}(1, z') = 1 - q$  despite the fact that  $z > z'$ . In other words, the contract paying more for completion of the task will produce worse beliefs about the task, conditional upon completion of the task. ■

This simple example serves to illustrate the inference of imperfectly recalled signals where there is no strategic interaction between the agent before the signal is forgotten and the agent after the signal has been forgotten. Since the ex-post beliefs are derived from the prior actions, we say that the agent engages in *self-justification*.

In this paper, we model an agent as illustrated in Example 1 who makes a decision regarding effort in two periods. We provide a definition of optimal behavior in the context of imperfect memory, which we refer to as Self-Justification Optimal (*SJO*). *SJO* specifies that at each point in which information has been learned or forgotten, the agent is modeled as a separate decision maker.

We provide a formal definition of a *Preference for Increasing Payments (PIP)*. This definition provides a standard such that if we observe *PIP* then we conclude that the agent has a genuine preference for increasing payments. The agent will consider a constant contract  $w^C$  (which pays the same amount for high effort in both periods) and an increasing contract  $w^I$  (which pays more for second period effort than first period effort). The definition of *PIP* is satisfied if  $w^C$  has a larger expected utility of the payments than  $w^I$ , the contracts are not too dissimilar and the agent prefers  $w^I$  over  $w^C$ . Therefore, if the conditions of *PIP* are satisfied then we can conclude that the choice could *only* have been made by a person with an intrinsic preference for an increasing sequence of payments.

To see how a preference for increasing payments can emerge, consider that the optimal first period strategy will condition on the first period signal. However, since this first period signal is to be forgotten, second period beliefs are a function of the strategies associated with each of the possible first period signals. Suppose we offer the agent a constant contract. Further, suppose that the optimal first period strategy given a signal of a low cost is to expend high effort with certainty and the first period strategy given a signal of a high cost is to mix: expend high effort with some probability and low effort with some probability. This mixture exchanges lower utility in the first period with higher utility in the second period, in the form of more favorable beliefs.

Now consider an increasing contract which makes the mixture probability under the constant contract more costly in the first period and no more beneficial in terms of second period beliefs. Hence, the agent will engage in this signaling less: by selecting a mixture which

exerts high effort with a lower probability than that given the constant contract. Therefore, when selecting a contract before any signals have been received, the agent might prefer an increasing contract over a constant contract, despite that the former has a smaller expected utility of payments than the latter.

## 1.1 Related Literature

Evidence related to the preference for increasing sequences of payments was first presented by Loewenstein and Sicherman (1991). The authors offered subjects a choice among of hypothetical payment profiles. Loewenstein and Sicherman found that a majority of the respondents preferred an increasing payment profile over a constant profile, despite that the constant profile had a larger present discounted value.<sup>2</sup> The authors also found that this preference for increasing payments is more pronounced when the payments are described as "income from wages" as opposed to "income from rent." Our paper addresses the second finding. We could easily modify the model to accommodate the evidence that people also have a preference for increasing payments of all types, however this would needlessly complicate the model. Therefore, in our model agents who expend an unknown, imperfectly recalled cost of effort (income from wages) can exhibit a preference for increasing payments and agents who expend a known, zero cost of effort (income from rent) cannot exhibit such a preference.

There exists a literature which models economic agents with imperfect memory. This concept has been used in a variety of applications, however to our knowledge, imperfect memory has not been used in connection with a preference for increasing payments.

We are not the first to assign a different status of memory on the basis of the type of information. For instance, Mullainathan (2002) makes a distinction between "hard" information and "soft" information where it is assumed that hard information is perfectly recalled and soft information is not. Like the present paper, Mullainathan considers private information to be soft and publicly available information to be hard. Other examples of authors assuming that actions and payments are not forgotten and private signals are forgotten include Swank (2006) and Hirshleifer and Welch (2002). Indeed we find evidence for this distinction in two separate strands of the psychology literature. In the first, researchers find that subjects remember the decisions themselves more than the reasons for those decisions. For instance, Arkes and Harkness (1980) find that subjects exhibited much better recall of their diagnosis than the symptoms which lead to the diagnosis.<sup>3</sup> In the second strand, described as the Availability Heuristic, researchers find a relationship between ease of recall and assessment of frequency and probability.<sup>4</sup> From this literature, we note that not only is the content of the information

---

<sup>2</sup>Later confirmed by Gigliotti and Sopher (1997), Guysu et. al. (2002) and Matsumoto et. al. (2000). See Gigliotti and Sopher (2003), Manzini et. al. (2006) and Read and Powell (2002) for other perspectives. See Clark (1999) and Grund and Sliwka (2007) for empirical evidence that increasing wages are associated with job satisfaction.

<sup>3</sup>This result was later confirmed by Lee and Uhlemann (1994).

<sup>4</sup>Tversky and Kahnemann (1973) are credited with starting this strand of literature. Also, see Gabrielcik and Fazio (1984) and Schwartz et. al. (1991).

relevant, but also the ease with which it is recalled.

We are not the first to model an agent with imperfect memory making inferences about personal characteristics through past actions. We are also not the first to assume that, subsequent to memory loss, the agent is aware that information as has been forgotten. In Benabou and Tirole (2004) a decision maker is both subject to imperfect recall and anticipates this imperfect recall. Benabou and Tirole assume that the agent is uncertain of the extent of their time inconsistency and makes an inference based on past actions. Benabou and Tirole assume that actions can be forgotten, by contrast we assume that only signals can be forgotten. Indeed the mechanism of making inferences of past characteristics through past actions in this manner was suggested by Benabou and Tirole (2003).<sup>5</sup> The novelty of this aspect of our modeling technique lies in analyzing its relationship with the preference for increasing payments.

Similar to Bernheim and Thomsen (2005), we model the agent with imperfect recall as relatively sophisticated and use a similar concept of optimality.<sup>6</sup> We assume that the agent can be modeled as several distinct players, each corresponding to the point in time in which their information is unique. In other words, every time information has been gained or lost, the agent is modeled as a separate decision maker. This implies that, in any period, the agent can only deviate from the optimal strategy among the strategies available to him during that particular period.

In this paper, it is assumed that the imperfect memory comes in the form that the agent recalls the fact that the information has been forgotten. By contrast, there is a literature which analyzes the issues which arise when such assumption is not made. Wilson (2004) models a decision problem of an agent with a limited memory capacity.<sup>7</sup> Wilson assumes that agents cannot determine the period in which each element of the memory capacity was obtained. Therefore, the agent does not know whether subsequent information has been forgotten. For papers which concentrate on the implications of this form of imperfect recall, without explicitly modeling its mechanism, see the literature on both the Absent Minded Driver Problem<sup>8</sup> and the Sleeping Beauty Problem.<sup>9</sup> This form of imperfect recall is less appropriate in our self-justification context. For self-justification, the agent must have *more* accurate recall of the action taken than of the private, personal the reasons for the action.

Benabou and Tirole (2003) present a model where higher payments can reduce an agent's utility from undertaking a task. In their model, an informed principal offers a contract to an uninformed agent. Specifically, the principal knows the cost of effort and the agent does not. The agent makes an inference regarding the cost of effort from the contract offered.

---

<sup>5</sup>Much in the spirit of Bem (1972).

<sup>6</sup>Which the authors refer to as "modified multiself consistency" and attribute to Piccione and Rubenstein (1997).

<sup>7</sup>See Dow (1991) for another model with a memory capacity.

<sup>8</sup>Piccione and Rubenstein (1997) along with the balance of the issue of Games and Economic Behavior.

<sup>9</sup>Elga (2000), Lewis (2001), Monton (2002), Dorr (2002), Bradley (2003) and Weintraub (2004).

By contrast, here we analyze a choice problem, as faced by the subjects in Loewenstein and Sicherman (1991), rather than a game between an informed principal and an uninformed agent.

The agent in our model is similar to an agent who has a preference for consistency because ex-post beliefs are updated to better conform to prior actions. Yariv (2006) models an agent with a preference for consistent beliefs by incorporating this as a separate term in the utility function. Therefore, the agent seeks a trade-off between material payoffs and consistency payoffs in selecting actions and beliefs. We distinguish between Yariv and this paper by noting that we do not assume a preference for consistency. Here the agent acts as if seeking to achieve consistency, but strives for the optimal action using the publicly available information in interpreting the past action. Also in this paper, the beliefs of the agent are rational in that they are not a subject of choice.

Eyster (2002) models an agent with a preference to avoid regret. This preference is incorporated into the utility function of the agent. The effort to avoid regret induces behavior as if the agent has a preference for consistency. A difference between Eyster and this paper is that Eyster requires actions to be ex-post suboptimal in order for the identified effect to occur. Specifically, all uncertainty must be resolved in order for the agent to determine if the action was regrettable. However, in this paper uncertainty is not resolved prior to the decision in either period. This should not be surprising as we are interested interpreting experiments in which a choice is made prior to the resolution of any uncertainty.

## 2 Model

We assume an agent with standard and separable preferences with regard to money. Utility for money ( $u : \mathbb{R}_+ \Rightarrow \mathbb{R}_+$ ) is everywhere increasing, concave and differentiable.

An agent is to complete a task in two periods. In periods 1 and 2, the agent decides to take the action or not ( $\{1, 0\} = A_t$  for  $t \in \{1, 2\}$ ). A contract<sup>10</sup> is a mapping from actions into payments ( $w_1 : A_1 \Rightarrow \mathbb{R}_+$  and  $w_2 : A_2 \Rightarrow \mathbb{R}_+$ ). For simplicity, we assume that  $w_t(0) = 0$ . Therefore, we can summarize the contract  $w$  by the pair of payments rendered for high effort ( $w_1, w_2$ ). Lower case  $w$  will denote a single contract and upper case  $W$  will denote a menu of contracts.

The cost of low effort is known to be 0. The cost of high effort is unknown. Priors regarding the cost of effort  $c$  are equally distributed on  $\{c_L, c_H\} = C$ . Further, we assume that  $c_H = 1$  and  $c_L = 0$ .

---

<sup>10</sup>Although we refer to these objects as 'contracts' this should not be interpreted as suggesting that we are analyzing strategic issues in a principal-agent setting. Here we exclusively focus on the choice problem of the agent as in Loewenstein and Sicherman (1991).

Before deciding the effort level in periods 1 and 2, the agent receives a signal  $s \in C$ , which imperfectly reveals the true cost of effort. We assume that:

$$P(s_j|c_i) = \begin{cases} q & \text{if } c_i = s_j \\ (1 - q) & \text{if } c_i \neq s_j \end{cases}$$

where  $q \in (0.5, 1)$ .<sup>11</sup> Conditional on  $c$ , we assume that the signals  $s_1$  and  $s_2$  are statistically independent. In the second period, the agent does not recall the first period signal and makes an inference regarding its content. In period 2, the agent will make this inference from all available information: the value of  $a_1$  (either 0 or 1) and the second period signal  $s_2$ .

The first and second period actions ( $a_1$  and  $a_2$ ) are mappings from all available information into a probability distribution on  $\{0, 1\}$ . The first period strategy can be written as  $a_1(s_1, w_1, w_2)$  and the second period strategy can be written as  $a_2(s_2, w_2, a_1(w_1, w_2))$ . Our notation reflects the above discussion as we show the dependence of  $a_2$  on the wages  $w_1, w_2$ , the action selected in the first period  $a_1$  and second period signal  $s_2$ , but not the first period signal  $s_1$ . Where there is no risk of confusion, we suppress the redundant notation in  $a_1(\cdot)$  and  $a_2(\cdot)$ . Also note that  $a_1(c_H)$  and  $a_1(c_L)$  (both  $\in [0, 1]$ ) refer to first period strategies and  $a_1 \in \{0, 1\}$  refers to the action actually taken.

To summarize the timing: in period 0 the agent selects a contract among those available. In periods 1 and 2 the agent determines whether to exert high or low effort given all information known during that period.

Now we provide the expected utility functions of the agent in each period. The expected utility of the agent in the second period, after receiving  $s_2$ , observing  $a_1$  and selecting  $a_2$ , can be written as:<sup>12</sup>

$$U_2(s_2, a_1) = a_2(s_2, a_1)\{u(w_2) - E_2[c|s_2, a_1]\}. \quad (1)$$

Here in expression (1), the second period agent can only observe the second period signal  $s_2$  and the value of  $a_1$ , as either 0 or 1, therefore we write  $E_2[c|s_2, a_1]$ .

The expected utility of the agent in the first period, after receiving  $s_1$  and selecting  $a_1$ , can be written as:

$$\begin{aligned} U_1(s_1) &= a_1(s_1)\{u(w_1) - E_1[c|s_1]\} \\ &+ \sum_{s_2 \in C} P(s_2|s_1)[a_1(s_1)a_2(s_2, a_1 = 1)\{u(w_2) - E_2[c|s_2, a_1 = 1]\} \\ &+ (1 - a_1(s_1))a_2(s_2, a_1 = 0)\{u(w_2) - E_2[c|s_2, a_1 = 0]\}]. \end{aligned} \quad (2)$$

Here in expression (2), the agent knows  $s_1$  and knows the function  $E_2[c|s_2, a_1]$  for each possible  $s_2$ . Furthermore, the agent in the first period uses  $s_1$  to improve the prediction of  $s_2$  through

<sup>11</sup>If  $q = 1$  then imperfect memory does not imply a preference for increasing payments.

<sup>12</sup>As these posterior beliefs are somewhat nonstandard, see the appendix for a more complete description.



$P(s_2|s_1)$ . It is worth noting that although the agent in the first period knows  $s_1$ , the agent also knows that the information will be forgotten in period 2. Therefore, sophistication requires that the first period consideration of second period utility should include  $E_2[c|s_2, a_1]$  rather than  $E_2[c|s_2, s_1]$ . If expression (2) contained a second period expectation which included  $s_1$ , the agent would be naive about the imperfect recall.

We write the expected utility of the contract  $w$  in the ex-ante period as:

$$\begin{aligned}
U_0(w) &= \frac{1}{2} \sum_{s_1 \in C} [a_1(s_1) \{u(w_1) - E_1[c|s_1]\}] \\
&+ \sum_{s_2 \in C} P(s_2|s_1) [a_1(s_1) a_2(s_2, a_1 = 1) \{u(w_2) - E_2[c|s_2, a_1 = 1]\}] \\
&+ (1 - a_1(s_1)) a_2(s_2, a_1 = 0) \{u(w_2) - E_2[c|s_2, a_1 = 0]\}.
\end{aligned} \tag{3}$$

Here in expression (3), the utility of the agent is the difference between the expected utility of money in each period and the expected cost of effort experienced in that period.

For convenience, in the upcoming definitions we denote the probability that the agent selects high effort in period  $t$  as  $\Psi_t$ . Therefore we can write

$$\Psi_1 = \frac{a_1^*(c_L) + a_2^*(c_H)}{2}$$

and

$$\begin{aligned}
\Psi_2 &= \frac{1}{2} \{a_2^*(c_L, a_1 = 1) (q a_1^*(c_L) + (1 - q) a_1^*(c_H)) \\
&+ a_2^*(c_L, a_1 = 0) (q(1 - a_1^*(c_L)) + (1 - q)(1 - a_1^*(c_H))) \\
&+ a_2^*(c_H, a_1 = 1) ((1 - q) a_1^*(c_L) + q a_1^*(c_H)) \\
&+ a_2^*(c_H, a_1 = 0) ((1 - q)(1 - a_1^*(c_L)) + q(1 - a_1^*(c_H)))\}.
\end{aligned}$$

We make two sophistication assumptions regarding the memory of the agent. We assume that the second period agent is sophisticated in remembering that the information has been forgotten. We also assume that the ex-ante and first period agents are sophisticated in that they anticipate this outcome. Although we regard these as strong assumptions, we think that some sophistication is reasonable. And while some sophistication is necessary for the results in this paper, we do not expect the results to qualitatively change if the agent is only partially sophisticated.

We now list our definition for optimal behavior: Self-Justification Optimal (*SJO*). Conditions (i), (ii) and (iii) require that at each period the agent maximize expected utility given what is known at the time. Specifically, we require that (Condition (i)) the ex-ante player selects the contract among the menu of contracts which will yield the highest expected utility. We require that (Condition (ii)) the first period agent maximizes expected utility given the

signal  $s_1$ . We require that (Condition (iii)) the second period agent maximizes expected utility given signal  $s_2$  and the behavior of the first period agent  $a_1$ . We also assume that when indifferent between selecting high and low effort in the second period, the agent selects low. We also make two additional requirements: a monotonicity requirement (Condition (M)) that the probability of effort induced by the low signal is at least as large that that from the high signal and a "reasonableness" requirement on the out-of-equilibrium beliefs (Condition (O)).

**Definition 1** *The agent is **Self-Justification Optimal (SJO)** if:*

- (i) *Contract  $w$  is selected where  $U_0(w) \geq U_0(w')$  for any  $w' \in W$ ,*
- (ii)  *$a_1^*(c_H)$  such that  $U_1(c_H, a_1^*(c_H)) \geq U_1(c_H, a_1(c_H))$  for any  $a_1(c_H)$  and  $a_1^*(c_L)$  such that  $U_1(c_L, a_1^*(c_L)) \geq U_1(c_L, a_1(c_L))$  for any  $a_1(c_L)$ ,*
- (iii)

$$a_2^*(s_2, a_1) = \begin{cases} 1 & \text{if } u(w_2) > E_2[c|s_2, a_1] \\ 0 & \text{if } u(w_2) \leq E_2[c|s_2, a_1] \end{cases} ,$$

(M)  $a_1^*(c_L) \geq a_1^*(c_H)$  and

(O) *out-of-equilibrium beliefs are such that if  $a_1^*(c_H) = a_1^*(c_L) = 1$  and  $a_1 = 0$  then period 2 agent infers that  $s_1 = c_H$  with probability one.*

We feel it worthwhile to describe the optimal behavior as *SJO* rather than a modified *PBE* so as to remind the reader of our imperfect memory setting and our related assumptions. Conditions (i), (ii) and (iii) constitute our optimality requirements for the agent in the ex-ante period, period 1 and period 2 respectively. Note that Condition (iii) is without loss of generality as any tie-breaking rule would not qualitatively change our results.

Condition (M) (or monotonicity) helps to eliminate from consideration, signaling outcomes which we consider to be inappropriate in the context of understanding the relationship between imperfect memory and the preference for increasing payments. For instance, a violation of condition (M) would imply that  $E[c|a_1 = 0] < E[c|a_1 = 1]$ . We wish to avoid such counterintuitive signaling outcomes. Condition (O) specifies the out-of-equilibrium beliefs after events with zero probability. It is assumed that if the first period strategy specifies that the agent always selects 1, however 0 is observed, then the agent infers that the agent received  $c_H$ .

Conditions (M) and (O) might strike the reader as rather blunt instruments. Alternatively, we could weaken these assumptions, however we would be obliged to consider optimal behavior which we do not consider to be helpful in understanding the relationship between imperfect memory and the preference for increasing payments.

We conclude this section with the following result which illustrates an important implication of optimality: an agent cannot mix at the same rate after both  $c_H$  and  $c_L$ . This result is significant as it captures the key insight that the first period agent exchanges current payoffs for an improvement in future beliefs. If there is no future benefit, as the inference after each

action is identical, then this exchange is not undertaken. The following proposition formalizes this statement.

**Proposition 1** *It cannot be an SJO that  $a_1^*(c_H) = a_1^*(c_L) \in (0, 1)$ .*

**Proof:** See appendix.

The intuition behind the result is as follows: the decision at period 1 is influenced by the anticipated posteriors in period 2. Particularly, if the first period agent selects an action other than the myopically optimal one then there must be some future benefit in the form of improved posteriors. If mixing occurs after both signals then it is certain that a myopically suboptimal first period action is taken. However, if the agent mixes at the same rate after both  $c_L$  and  $c_H$  then the posteriors are unaffected and so this action will not produce a benefit. Therefore, identical mixing after  $c_L$  and  $c_H$  is inconsistent with optimality.

### 3 A Preference for Increasing Payments

In this section, we discuss the main implication of the model: an agent with imperfect memory can display a preference for an increasing sequence of payments. To make this notion more precise we offer the definition of a *Preference for Increasing Payments (PIP)*.

Consider a constant contract  $w^C$  where  $w_1^C = w_2^C$ . *PIP* places restrictions on a candidate increasing contract  $w^I$  where  $w_1^I < w_2^I$  such that if  $w^I$  is preferred over  $w^C$  then we are able to conclude that this choice is the result of an intrinsic preference for increasing payments. Operationally, *PIP* requires that the candidate increasing contract  $w^I$  is less *valuable* than the constant contract  $w^C$ . Additionally,  $w^I$  must be *close* to  $w^C$  in that both contracts induce qualitatively similar behavior. If  $w^I$  satisfies these requirements of *valuableness* and *closeness* with respect to a  $w^C$  and additionally  $w^I$  is preferred over  $w^C$  then we conclude that the agent has an intrinsic preference for increasing payments.

Suppose that an agent is deciding between a constant contract  $w^C$  and an increasing contract  $w^I$ . If the expected utility of the payments from  $w^I$  exceeds the expected utility of payments from  $w^C$ , and the agent prefers  $w^I$ , we would be wrong to conclude that the agent exhibits a preference for increasing payments. The agent simply prefers more money to less. Therefore, in determining whether an agent exhibits a preference for increasing payments, we require that the expected utility of payments from  $w^C$  exceeds that from  $w^I$ . We refer to this requirement as  $w^I$  being less *valuable* than  $w^C$ .

**Definition 2** *A contract  $w^I$  is less **valuable** than contract  $w^C$  if*

$$\begin{aligned} & u(w_1^C)\Psi_1^C + u(w_2^C)\Psi_2^C \\ > & u(w_1^I)\Psi_1^I + u(w_2^I)\Psi_2^I. \end{aligned} \tag{4}$$

The definition of valuable is rather involved because the actions which comprise each  $\Psi$  behave in a nonobvious manner. However, we want to compare the expected utility of payments between two contracts therefore we have to consider the behavior induced by both contracts.

Now suppose that  $w^I$  is less valuable than  $w^C$ . However, also suppose that  $w^C$  and  $w^I$  are sufficiently dissimilar so that qualitatively different behavior is induced. Again, we would be wrong to conclude that the agent exhibits a preference for increasing payments. We would only be justified in concluding that the agent likes the overall characteristics of  $w^I$  more than those of  $w^C$ . In particular, we require the optimal second period strategies for  $w^I$  and  $w^C$  are identical. We also require that the optimal first period strategies for  $w^I$  and  $w^C$  are qualitatively similar. Specifically, we require strategies given  $s_1$  to be pure strategies of both 0, both 1 or both mixed. We refer to this requirement as  $w^I$  being *close* to  $w^C$ .

**Definition 3** *Contracts  $w^C$  and  $w^I$  are **close** if*

$$\begin{aligned} & \text{for } s_2 \in \{c_L, c_H\} \text{ either} & (5) \\ & a_1^{*C}(s_1) = a_1^{*I}(s_1) \in \{0, 1\} \text{ or} \\ & a_1^{*C}(s_1), a_1^{*I}(s_1) \in (0, 1) \end{aligned}$$

and

$$a_2^{*C}(s_2, a_1) = a_2^{*I}(s_2, a_1) \text{ for } s_2 \in \{c_L, c_H\} \text{ and } a_1 \in \{0, 1\}. \quad (6)$$

We now state the definition which provides the criteria for determining whether the agent displays a preference for increasing payments. We require that for such a determination, there must exist an increasing contract which is less *valuable* than a constant contract and that the two contracts are *close*. If these two conditions hold and the increasing contract is preferred over the constant contract then we declare that the agent exhibits a preference for increasing payments.

**Definition 4** *Consider a constant contract  $w^C$ . An SJO agent displays a **Preference for Increasing Payments (PIP)** if there exists an increasing contract  $w^I$  which is less valuable than  $w^C$ , close to  $w^C$  and*

$$U_0^*(w^I) > U_0^*(w^C).$$

Definition 4 provides a standard for determining whether the agent exhibits a preference for increasing payments. Intuitively, if contract  $w^C$  pays more than contract  $w^I$  and  $w^C$  is similar to  $w^I$  yet the agent prefers  $w^I$  to  $w^C$  then we say that the agent has an intrinsic preference for increasing payments. To better understand the content of Definition 4 we provide the following example.

**Example 2** *Assume that the agent perfectly recalls the signal and that  $q = 0.7$ . Contract  $w^C$  pays  $u(w_1^C) = u(w_2^C) = 0.5$ . Contract  $w^I$  pays  $u(w_1^I) = 0.49$  and  $u(w_2^I) = 0.51$ . Under*

$w^C$  the agent selects  $a_1 = 1$  only after  $s_1 = c_L$  and  $a_2 = 1$  only after  $s_2 = c_L$  and  $s_1 = c_L$ . Therefore, total utility from contract  $w^C$  is then

$$U_0(w^C) = 0.5[0.5 - 0.3 + 0.7[0.5 - \frac{0.09}{0.58}]] = 0.221.$$

Under  $w^I$  the agent again selects high effort only after  $s_1 = c_L$  however in the second period selects high effort for every signal pair other than  $s_2 = c_H$  and  $s_1 = c_H$ . Therefore the total utility from contract  $w^I$  is:

$$U_0(w^I) = 0.5[0.49 - 0.3 + 0.7[0.51 - \frac{0.09}{0.58}]] + 0.3[0.51 - 0.5] + 0.5[0.3[0.51 - 0.5]] = 0.222.$$

■

A few aspects of Example 2 are worth noting. The first is that the contracts are not "close" despite that they appear to be rather similar. However, the small difference between the two contracts leads to a large difference in the induced behavior. Our definition of close seeks to avoid these types of examples. The second notable aspect of the example relates to the assumption of perfect recall. In what follows, we show that an agent with perfect recall cannot exhibit *PIP* (Proposition 5).

## 4 *SJO* Behavior

We now provide an explicit characterization of the relationship between the parameter values and the exhibition of *PIP*.<sup>13</sup> Together, the propositions below yield a novel implication of our model: an agent displays *PIP* when it is not the case that the payments are very likely to cover the cost of high effort and it is not the case that the payments are very unlikely to cover the cost of high effort. In these cases, the agent will find it worthwhile to seek a reduction in the perceived cost of effort by accepting a close and less valuable contract which induces more favorable beliefs.

We start out with two negative results. The content of these two propositions can be summarized by the following: anytime behavior is identical between two close contracts, the more valuable one will be preferred.<sup>14</sup> Contracts in the parameter values described below will always induce identical behavior for  $w^C$  and any close  $w^I$ .

**Proposition 2** *If (i)  $u(w_1^C) = u(w_2^C) \in (1 - q, \frac{1}{2})$  or (ii)  $u(w_1^C) = u(w_2^C) > \frac{q^2}{q^2 + (1-q)^2}$  then the agent never exhibits *PIP*.*

**Proof:** See appendix.

<sup>13</sup>For a complete characterization of *SJO*, see Lemmas 4 through 7 in the appendix.

<sup>14</sup>Indeed, this is the content of Proposition 6 in the following section.

**Proposition 3** (i) If  $u(w_1^C) = u(w_2^C) \in (\frac{1}{2}, q)$  then there is a  $\lambda^* \in (\frac{1}{2}, q)$  such that if  $u(w_1^C) = u(w_2^C) \in (\frac{1}{2}, \lambda^*)$  then the agent never exhibits *PIP*.

(ii) If  $u(w_1^C) = u(w_2^C) \in (q, \frac{q^2}{q^2+(1-q)^2})$  then there is a  $\mu^* \in (q, \frac{q^2}{q^2+(1-q)^2})$  such that if  $u(w_1^C) = u(w_2^C) \in (\mu^*, \frac{q^2}{q^2+(1-q)^2})$  then the agent never exhibits *PIP*.

**Proof:** See appendix.

The intuition is that for these parameter values, behavior is identical for  $w^C$  and any close  $w^I$  therefore (see Proposition 6) these contracts cannot constitute a *PIP*. Note the parameter values in the above propositions. There are two possibilities: either it is very unlikely that the contract will cover a high cost of effort (Propositions 2 (i) and 3 (i)) or it is very likely that the contract will cover a high cost of effort (Propositions 2 (ii) and 3 (ii)). In these cases, the agent cannot satisfy *PIP*.

Although the above two results are negative, we now provide a positive result. The following proposition states that any time an agent is considering a contract  $w^C$  where  $a_1^{*C}(c_H) \in (0, 1)$  and  $a_1^{*C}(c_L) = 1$  then it is always possible to find a less valuable and close increasing contract which is preferred over the constant contract. In other words, Proposition 4 says that the agent always displays *PIP* when  $a_1^{*C}(c_H) \in (0, 1)$  and  $a_1^{*C}(c_L) = 1$ .

**Proposition 4** (i) If  $u(w_1^C) = u(w_2^C) \in (\frac{1}{2}, q)$  then there is a  $\lambda^* \in (\frac{1}{2}, q)$  such that if  $u(w_1^C) = u(w_2^C) \in (\lambda^*, q)$  then the agent always exhibits *PIP*.

(ii) If  $u(w_1^C) = u(w_2^C) \in (q, \frac{q^2}{q^2+(1-q)^2})$  then there is a  $\mu^* \in [q, \frac{q^2}{q^2+(1-q)^2})$  such that if  $q < \mu^*$  and  $u(w_1^C) = u(w_2^C) \in (q, \mu^*)$  then the agent always exhibits *PIP*.

**Proof:** See appendix.

Although the proof of Proposition 4 is involved, the intuition is straightforward. Anytime  $a_1^{*C}(c_H) \in (0, 1)$  and  $a_1^{*C}(c_L) = 1$ , the ex-ante player prefers the first period player to select a smaller  $a_1(c_H)$  than is preferred by the first period agent. So for these parameter values, it is always possible to find a close and less valuable increasing contract which induces a sufficiently smaller  $a_1^{*C}(c_H)$  in order to more than compensate for the smaller expected utility of payments of the increasing contract. The agent selects a smaller  $a_1(c_H)$  under the increasing contract because the marginal cost of the influencing second period beliefs is higher than under the constant contract, therefore the agent engages is it less. Therefore, the agent satisfies *PIP*. Note the different wording for parts (i) and (ii) of Proposition 4. This difference is required as there exist parameter values such that  $u(w_1^C) = u(w_2^C)$  and  $q = \mu^*$ .

Roughly, Propositions 2, 3 and 4 say that if payment for the task neither very likely nor very unlikely to cover a high cost of effort, then the agent will exhibit *PIP*. This is a novel implication of our model.

As previously mentioned, the effects just cited only occur when the agent has imperfect memory. Indeed this is the content of the following proposition.

**Proposition 5** *If an agent has perfect recall of the signal of the cost of effort then the agent cannot display PIP.*

**Proof:** See appendix.

The intuition for the result is as follows. With perfect memory, first period actions do not affect second period beliefs. As a result, the first period action is not selected in order to influence second period beliefs. Additionally, when the valuable condition is met then it must be that  $U_0(w^C) - U_0(w^I) < 0$ . Therefore, the agent with perfect memory does not display PIP. Although the result appears to be straightforward, we hope to convince the reader of its significance as Proposition 5 supports our proposal of modeling the preference for increasing payments through imperfect memory.

To give the reader some intuition regarding PIP we provide the following example.

**Example 3** *Suppose that  $q = 0.7$ . The agent is considering two contracts:  $w^C$  and  $w^I$ . Contract  $w^C$  pays  $u(w^C) = 0.6$ . Contract  $w^I$  pays  $u(w_1^I) = 0.59$  and  $u(w_2^I) = 0.609$ .<sup>15</sup> It follows that for both contracts*

$$\begin{aligned} a_2^*(c_L, a_1 = 1) &= a_2^*(c_L, a_1 = 0) = 1 \\ a_2^*(c_H, a_1 = 0) &= 0. \end{aligned}$$

*It is SJO under both contracts that  $a_1^*(c_L) = 1$  for every choice of  $a_1(c_H)$ . The first period player receiving  $s_1 = c_H$  will seek to maximize:*

$$\begin{aligned} U_1(c_H) &= a_1(c_H)\{u(w_1) - 0.7\} \\ &+ 0.7a_1(c_H)a_2(s_H, a_1 = 1)\{u(w_2) - E[c|c_H, a_1 = 1]\} \\ &+ 0.3a_1(c_H)\{u(w_2) - E[c|c_L, a_1 = 1]\} \\ &+ 0.3(1 - a_1(c_H))\{u(w_2) - E[c|c_L, a_1 = 0]\}. \end{aligned}$$

*Contract  $w^C$  induces  $a_1^{*C}(c_H) = 0.109$  and  $a_2^{*C}(s_H, a_1 = 1) = 1$ . This implies that  $U_1^*(w^C, c_H) = 0.0338$ ,  $U_1^*(w^C, c_L) = 0.610$  and an ex-ante utility of  $2U_0^*(w^C) = 0.644$ . Contract  $w^I$  induces  $a_1^{*I}(c_H) = 0.102$  and  $a_2^{*I}(s_H, a_1 = 1) = 1$ . This implies that  $U_1^*(w^I, c_H) = 0.036$ ,  $U_1^*(w^I, c_L) = 0.611$  and an ex-ante utility of  $2U_0^*(w^I) = 0.647$ . Therefore, the ex-ante player prefers  $w^I$  to  $w^C$ . To see that the valuable requirement is satisfied:*

$$\begin{aligned} &0.6(0.5)(1.0109) + 0.6[1 - (0.5)(0.7)(1 - 0.109)] \\ &> 0.59(0.5)(1.102) + 0.607[1 - (0.5)(0.7)(1 - 0.102)]. \end{aligned} \tag{7}$$

*Note also that close requirement is satisfied. Therefore, the agent exhibits PIP. ■*

---

<sup>15</sup>Suppose that  $u(w) = w^{\frac{2}{3}}$ . Therefore,  $w_1^C = w_2^C = 0.465$ ,  $w_1^I = 0.453$ ,  $w_2^I = 0.475$  and  $w_1^C + w_2^C > w_1^I + w_2^I$ . In other words here again the agent actually prefers a smaller total payment so long as it is increasing.

The above example provides intuition behind the definitions of valuable and close. In Example 3, the contracts are such that  $u(w_1^I) = 0.59$   $u(w_2^I) = 0.609$  and  $u(w_1^C) = u(w_2^C) = 0.6$ . Although it seems natural to conclude that the pairs are similar, one can confirm that the requirements for close are satisfied: both induce identical second period strategies and both have qualitatively similar first period strategies. Expression (7) demonstrates valuableness which requires that the expected utility of payment from  $w^C$  is greater than the expected utility of payment from  $w^I$ .

We now provide a proposition which states that the exhibition of *PIP* is only generated by behavior.

**Proposition 6** *Consider contracts  $w^I$  and  $w^C$ . If first period behavior is identical then *PIP* cannot occur.*

**Proof:** See appendix.

The intuition behind Proposition 6 is as follows: the exhibition of *PIP* requires different posteriors for the two contracts and affecting the posteriors can only be achieved through the application different first period strategies for each contract. If the contracts do not induce such behavior then an increasing contract will never fulfill the requirements in *PIP*. Apart from providing intuition, Proposition 6 also has practical relevance. Suppose that we have identified the optimal strategies for two contracts. If the strategies are identical then the contract cannot be part of an exhibition of *PIP*.

## 5 Conclusion

In this paper we have analyzed the role of imperfect memory in the preference for increasing payments. Our model aids in the interpretation of existing choice experiments which suggest that people have a stronger preference for increasing sequences of payments when described as "income from wages" rather than "income from rent."

With the auxiliary assumption that there is an unknown cost of effort in obtaining the former and not the latter, our model provides a mechanism through which this behavior occurs. Our model not only replicates the Loewenstein and Sicherman results but also makes the additional prediction that such behavior will only be associated with contracts which are neither very likely nor very unlikely to cover the cost of high effort. There are other explanations for the difference between the preferences of increasing wages and rent. However, these explanations require a distinction between the two which we believe to be less natural than ours: income from wages requires the repeated expenditure of unknown, imperfectly recalled cost of effort and income from rent requires a known, zero cost of effort.

It should be noted that the results of this paper would not change if we were to give the agent a pencil and paper (or a similar means of remembering the first period signal). In the



first period, each of the possible signals induces different incentives: the low signal recipient wishes to truthfully reveal the signal and the high signal recipient wishes to conceal the true signal. As a result, the high signal recipient would simply indicate on the paper that the signal is low, and the results would remain unchanged.

There are several issues which deserve further scrutiny. For instance, we do not know the significance of our choice of state space. It would be interesting to learn the *SJO* behavior of our agent with a discrete state space of more than two elements or a continuous state space. It would also be interesting to learn the behavior of the agent if there are more than two decisions regarding effort. Finally, it would be interesting to learn the weight of our sophistication requirements. Although we will need some sophistication for the results to hold, future work will hopefully specify the relationship between these weaker assumptions and behavior.

## 6 Appendix

The appendix is arranged as follows. In the first subsection, we first derive the posterior beliefs of the agent. These require attention as they are somewhat non-standard. Then we provide a more complete description of the utilities of the agent. In the following subsection, we prove the lemmas which will be useful later. In the final subsection, we prove the results found in the body of the paper. We present the proofs of the results in the order to best facilitate their elucidation rather than the order presented in the body of the paper.

### 6.1 Preliminaries

We now derive the posterior beliefs of the period 2 agent given  $s_2$  and  $a_1 = 1$  as found in expression (1). If it is not the case that either  $a_1(c_L) = a_1(c_H) = 1$  or  $a_1(c_L) = a_1(c_H) = 0$  then

$$P(\hat{c}|s_2, a_1) = \frac{P(s_2|\hat{c})P(a_1|\hat{c})}{\sum_{c \in C} P(s_2|c)P(a_1|c)}$$

since  $s_1$  and  $s_2$  are independent conditional on  $c$ . More explicitly

$$P(c|s_2 = c_H, a_1 = 1) = \begin{cases} \frac{q(qa_1(c_H) + (1-q)a_1(c_L))}{(q^2 + (1-q)^2)a_1(c_H) + 2q(1-q)a_1(c_L)} & \text{if } c = c_H \\ \frac{(1-q)((1-q)a_1(c_H) + qa_1(c_L))}{(q^2 + (1-q)^2)a_1(c_H) + 2q(1-q)a_1(c_L)} & \text{if } c = c_L \end{cases} \quad (8)$$

$$P(c|s_2 = c_L, a_1 = 1) = \begin{cases} \frac{(1-q)(qa_1(c_H) + (1-q)a_1(c_L))}{2q(1-q)a_1(c_H) + (q^2 + (1-q)^2)a_1(c_L)} & \text{if } c = c_H \\ \frac{q((1-q)a_1(c_H) + qa_1(c_L))}{2q(1-q)a_1(c_H) + (q^2 + (1-q)^2)a_1(c_L)} & \text{if } c = c_L \end{cases} \quad (9)$$

and

$$P(c|s_2 = c_H, a_1 = 0) = \begin{cases} \frac{q(q(1-a_1(c_H)) + (1-q)(1-a_1(c_L)))}{(q^2 + (1-q)^2)(1-a_1(c_H)) + 2q(1-q)(1-a_1(c_L))} & \text{if } c = c_H \\ \frac{(1-q)((1-q)(1-a_1(c_H)) + q(1-a_1(c_L)))}{(q^2 + (1-q)^2)(1-a_1(c_H)) + 2q(1-q)(1-a_1(c_L))} & \text{if } c = c_L \end{cases} \quad (10)$$

$$P(c|s_2 = c_L, a_1 = 0) = \begin{cases} \frac{(1-q)(q(1-a_1(c_H)) + (1-q)(1-a_1(c_L)))}{2q(1-q)(1-a_1(c_H)) + (q^2 + (1-q)^2)(1-a_1(c_L))} & \text{if } c = c_H \\ \frac{q((1-q)(1-a_1(c_H)) + q(1-a_1(c_L)))}{2q(1-q)(1-a_1(c_H)) + (q^2 + (1-q)^2)(1-a_1(c_L))} & \text{if } c = c_L \end{cases} \quad (11)$$

We can write the expectation as:

$$E[c|s_2, a_1] = c_L P(c_L|s_2, a_1) + c_H P(c_H|s_2, a_1)$$

and since we have defined  $c_H = 1$  and  $c_L = 0$  we can write:

$$E[c|s_2, a_1] = P(c_H|s_2, a_1).$$

If it is not the case that either  $a_1(c_H) = a_1(c_L) = 1$  or  $a_1(c_H) = a_1(c_L) = 0$  then we can write expression (2) conditional on  $c_L$  as:

$$\begin{aligned}
U_1(c_L) &= a_1(c_L)(u(w_1) - (1 - q)) \tag{12} \\
&+ (1 - q)a_1(c_L)a_2(c_H, a_1 = 1)(u(w_2) - \left( \frac{q^2 a_1(c_H) + q(1 - q)a_1(c_L)}{(q^2 + (1 - q)^2)a_1(c_H) + 2q(1 - q)a_1(c_L)} \right)) \\
&\quad + (1 - q)(1 - a_1(c_L))a_2(c_H, a_1 = 0)(u(w_2) - \\
&\quad \left( \frac{q^2(1 - a_1(c_H)) + q(1 - q)(1 - a_1(c_L))}{(q^2 + (1 - q)^2)(1 - a_1(c_H)) + 2q(1 - q)(1 - a_1(c_L))} \right)) \\
&\quad + qa_1(c_L)a_2(c_L, a_1 = 1)(u(w_2) - \left( \frac{q(1 - q)a_1(c_H) + (1 - q)^2 a_1(c_L)}{2q(1 - q)a_1(c_H) + (q^2 + (1 - q)^2)a_1(c_L)} \right)) \\
&+ q(1 - a_1(c_L))a_2(c_L, a_1 = 0)(u(w_2) - \left( \frac{q(1 - q)(1 - a_1(c_H)) + (1 - q)^2(1 - a_1(c_L))}{2q(1 - q)(1 - a_1(c_H)) + (q^2 + (1 - q)^2)(1 - a_1(c_L))} \right))
\end{aligned}$$

and the analogous expression for  $U_1(c_H)$ .

## 6.2 Supporting Results

Lemma 1 simplifies the decision problem as, out of the four second period actions, the value of at most one is determined by first period strategies.

**Lemma 1** *Given Conditions M and O in Definition 1:*

$$\begin{aligned}
\frac{q^2}{q^2 + (1 - q)^2} &\geq E_2[c|s_H, a_1 = 0] \geq q \geq E_2[c|s_H, a_1 = 1] \geq 0.5 \\
&\geq E_2[c|s_L, a_1 = 0] \geq 1 - q \geq E_2[c|s_H, a_1 = 1] \geq \frac{(1 - q)^2}{q^2 + (1 - q)^2}.
\end{aligned}$$

**Proof of Lemma 1:** When it is not the case that either  $a_1(c_H) = a_1(c_L) = 1$  or  $a_1(c_H) = a_1(c_L) = 0$

$$\begin{aligned}
E[c|s_H, a_1 = 0] &= \left( \frac{q^2(1 - a_1(c_H)) + q(1 - q)(1 - a_1(c_L))}{(q^2 + (1 - q)^2)(1 - a_1(c_H)) + 2q(1 - q)(1 - a_1(c_L))} \right) \\
E[c|s_H, a_1 = 1] &= \left( \frac{q^2 a_1(c_H) + q(1 - q)a_1(c_L)}{(q^2 + (1 - q)^2)a_1(c_H) + 2q(1 - q)a_1(c_L)} \right) \\
E[c|s_L, a_1 = 0] &= \left( \frac{q(1 - q)(1 - a_1(c_H)) + (1 - q)^2(1 - a_1(c_L))}{2q(1 - q)(1 - a_1(c_H)) + (q^2 + (1 - q)^2)(1 - a_1(c_L))} \right) \\
E[c|s_L, a_1 = 1] &= \left( \frac{q(1 - q)a_1(c_H) + (1 - q)^2 a_1(c_L)}{2q(1 - q)a_1(c_H) + (q^2 + (1 - q)^2)a_1(c_L)} \right).
\end{aligned}$$

Given condition  $M$  it must be that:

$$\begin{aligned} \frac{q^2}{q^2 + (1-q)^2} &\geq E[c|s_H, a_1 = 0] \geq q \\ q &\geq E[c|s_H, a_1 = 1] \geq 0.5 \\ 0.5 &\geq E[c|s_L, a_1 = 0] \geq 1-q \\ 1-q &\geq E[c|s_L, a_1 = 1] \geq \frac{(1-q)^2}{q^2 + (1-q)^2}. \end{aligned}$$

If  $a_1(c_H) = a_1(c_L) = 1$  then given condition  $O$ :

$$\begin{aligned} E[c|s_H, a_1 = 0] &= \frac{q^2}{q^2 + (1-q)^2} \\ E[c|s_H, a_1 = 1] &= q \\ E[c|s_L, a_1 = 0] &= 0.5 \\ E[c|s_L, a_1 = 1] &= 1-q. \end{aligned}$$

If  $a_1(c_H) = a_1(c_L) = 0$  then given condition  $O$ :

$$\begin{aligned} E[c|s_H, a_1 = 0] &= q \\ E[c|s_H, a_1 = 1] &= 0.5 \\ E[c|s_L, a_1 = 0] &= 1-q \\ E[c|s_L, a_1 = 1] &= \frac{(1-q)^2}{q^2 + (1-q)^2}. \end{aligned}$$

■

**Lemma 2** *If  $u(w_1) > 1-q$  then*

$$a_1(c_L) = 1$$

*will be an SJO best response to every  $a_1(c_H) \in [0, 1]$ .*

**Proof of Lemma:** Suppose that  $a_1(c_H) = 1$  then  $a_1(c_L) = 1$  by Condition  $M$ .

Suppose that  $a_1(c_H) = 0$ . In this case for  $a_1(c_L) > 0$ :

$$\begin{aligned} U_1(c_L) &= a_1(c_L)(u(w_1) - (1-q)) + (1-q)a_1(c_L)a_2(c_H, a_1 = 1)(u(w_2) - \frac{1}{2}) \quad (13) \\ &+ (1-q)(1-a_1(c_L))a_2(c_H, a_1 = 0)(u(w_2) - \left( \frac{q^2 + q(1-q)(1-a_1(c_L))}{(q^2 + (1-q)^2) + 2q(1-q)(1-a_1(c_L))} \right)) \\ &\quad + qa_1(c_L)a_2(c_L, a_1 = 1)(u(w_2) - \left( \frac{(1-q)^2}{q^2 + (1-q)^2} \right)) \\ &+ q(1-a_1(c_L))a_2(c_L, a_1 = 0)(u(w_2) - \left( \frac{q(1-q) + (1-q)^2(1-a_1(c_L))}{2q(1-q) + (q^2 + (1-q)^2)(1-a_1(c_L))} \right)). \end{aligned}$$

Since  $\frac{1}{2} < \frac{q^2+q(1-q)(1-a_1(c_L))}{(q^2+(1-q)^2)+2q(1-q)(1-a_1(c_L))}$  and  $\frac{(1-q)^2}{q^2+(1-q)^2} < \frac{q(1-q)+(1-q)^2(1-a_1(c_L))}{2q(1-q)+(q^2+(1-q)^2)(1-a_1(c_L))}$  for every  $a_1(c_L)$ , and  $u(w_1) > 1 - q$ ,  $U_1(c_L)$  is an increasing function of  $a_1(c_L)$  with a maximum at 1. Therefore for  $a_1(c_L) > 0$ ,  $a_1(c_L) = 1$  is a best response to  $a_1(c_H) = 0$ .

In the case that  $a_1(c_L) = 0$ :

$$U_1(c_L) = (1 - q)a_2(c_H, a_1 = 0)(u(w_2) - q) + a_2(c_L, a_1 = 0)(u(w_2) - (1 - q)). \quad (14)$$

Selecting  $a_1(c_L) > 0$  does strictly better than  $a_1(c_L) = 0$  as expression (13) is always greater than expression (14). And so if  $a_1(c_H) = 0$  then  $a_1(c_L) = 1$  is a best response.

Suppose that  $a_1(c_H) \in (0, 1)$  then expression (12) applies. Just as in the  $a_1(c_H) = 0$  case  $U_1(c_L)$  is an increasing function of  $a_1(c_L)$  (with the domain of  $a_1(c_L) \in [0, 1]$ ) with a maximum at 1 and the Lemma is proved.

■

**Lemma 3** *Suppose that  $a, a', b, b', c, c' > 0$ . If  $1 \geq y > x \geq 0$  and  $\frac{a}{a'} \geq \frac{b}{b'} \geq \frac{c}{c'}$ , with at least one inequality strict then*

$$\frac{ay^2 + by + c}{a'y^2 + b'y + c'} > \frac{ax^2 + bx + c}{a'x^2 + b'x + c'}. \quad (15)$$

**Proof:** We rewrite the expression (15) as:

$$\begin{aligned} & ab'xy^2 + ac'y^2 + a'bx^2y + bc'y + a'cx^2 + b'cx \\ & > a'bx^2y + a'cy^2 + ab'x^2y + b'cy + ac'x^2 + bc'x \end{aligned}$$

which is equivalent to:

$$xy(y - x)(ab' - a'b) + (y^2 - x^2)(ac' - a'c) + (y - x)(bc' - b'c) > 0.$$

The above expression always holds when  $\frac{a}{a'} \geq \frac{b}{b'} \geq \frac{c}{c'}$  and when at least one inequality holds strictly. Therefore, the lemma is proved.

■

**Lemma 4** *If  $u(w_1), u(w_2) \in (1 - q, \frac{1}{2}]$  then it is SJO that  $a_1^*(c_H) = 0$  and  $a_1^*(c_L) = 1$ .*

**Proof of Lemma 4:** By Lemma 2, it must be that  $a_1^*(c_L) = 1$  which implies that  $E[c|c_L, a_1 = 0] = \frac{1}{2}$ . Therefore, by assumption  $u(w_2) \leq \frac{1}{2}$  it must be that  $a_2^*(c_L, a_1 = 0) = 0$ . Since  $u(w_2) > 1 - q$  it will be that  $a_2^*(c_L, a_1 = 1) = 1$ . We can write

$$\begin{aligned} U_1(c_H) &= a_1(c_H)(u(w_1) - q) \\ &+ (1 - q)a_1(c_H)(u(w_2) - \left( \frac{q(1 - q)a_1(c_H) + (1 - q)^2}{2q(1 - q)a_1(c_H) + (q^2 + (1 - q)^2)} \right)). \end{aligned}$$

Therefore,

$$\frac{\partial U_1(c_H)}{\partial a_1(c_H)} = u(w_1) - q$$

$$+(1-q)(u(w_2) - \left( \frac{2q^2(1-q)^2 a_1(c_H)^2 + 2q(1-q)(q^2 + (1-q)^2) a_1(c_H) + (1-q)^2(q^2 + (1-q)^2)}{4q^2(1-q)^2 a_1(c_H)^2 + 4q(1-q)(q^2 + (1-q)^2) a_1(c_H) + (q^2 + (1-q)^2)^2} \right)).$$

By Lemma 3,  $\frac{\partial U_1(c_H)}{\partial a_1(c_H)}$  is everywhere decreasing in  $a_1(c_H)$ . Additionally,  $\frac{\partial U_1(c_H)}{\partial a_1(c_H)}$  is everywhere strictly less than zero as

$$\frac{1}{2} - q + (1-q) \left( \frac{1}{2} - \frac{(1-q)^2}{q^2 + (1-q)^2} \right) = \frac{1}{2} - q + (1-q) \left( \frac{q - \frac{1}{2}}{q^2 + (1-q)^2} \right)$$

which is less than zero and so it must be that  $a_1^*(c_H) = 0$ .

■

**Lemma 5** For every  $u(w_2) \in (\frac{1}{2}, q)$  there exists a  $\lambda^*(u(w_2)) \in [\frac{1}{2}, q)$  such that for all  $u(w_1) \in (\frac{1}{2}, \lambda^*(u(w_2)))$  it is SJO that  $a_1^*(c_H) = 0$  and  $a_1^*(c_L) = 1$  and for all  $u(w_1) \in (\lambda^*(u(w_2)), q)$  it is SJO that  $a_1^*(c_H) \in (0, 1)$  and  $a_1^*(c_L) = 1$ . Additionally, if  $u(w_1) = u(w_2)$  then  $\lambda^* > \frac{1}{2}$  and  $a_2^*(c_H, a_1 = 1) = 1$

**Proof:** By Lemma 2 it must be that  $a_1^*(c_L) = 1$ . Since  $u(w_1), u(w_2) \in (\frac{1}{2}, q)$  it will be that

$$a_2^*(c_L, a_1 = 0) = a_2^*(c_L, a_1 = 1) = 1$$

$$a_2^*(c_H, a_1 = 0) = 0.$$

We can write:

$$U_1(c_H) = a_1(c_H)(u(w_1) - q) \tag{16}$$

$$+(1-q)a_1(c_H)(u(w_2) - \left( \frac{q(1-q)a_1(c_H) + (1-q)^2}{2q(1-q)a_1(c_H) + (q^2 + (1-q)^2)} \right))$$

$$+(1-q)(1-a_1(c_H))(u(w_2) - \frac{1}{2})$$

$$+qa_1(c_H)a_2(c_H, a_1 = 1)(u(w_2) - \left( \frac{q^2 a_1(c_H) + q(1-q)}{(q^2 + (1-q)^2)a_1(c_H) + 2q(1-q)} \right))$$

and therefore

$$\frac{\partial U_1(c_H)}{\partial a_1(c_H)} = u(w_1) - q \tag{17}$$

$$+(1-q)(0.5 - \left[ \frac{2q^2(1-q)^2 a_1(c_H)^2 + 2q(1-q)(q^2 + (1-q)^2) a_1(c_H) + (1-q)^2(q^2 + (1-q)^2)}{4q^2(1-q)^2 a_1(c_H)^2 + 4q(1-q)(q^2 + (1-q)^2) a_1(c_H) + (q^2 + (1-q)^2)^2} \right])$$

$$+qa_2(c_H, a_1 = 1)(u(w_2) - \left[ \frac{q^2(q^2 + (1-q)^2) a_1(c_H)^2 + 4q^3(1-q) a_1(c_H) + 2q^2(1-q)^2}{(q^2 + (1-q)^2)^2 a_1(c_H)^2 + 4q(1-q) a_1(c_H) + 4q^2(1-q)^2} \right]).$$

By Lemma 3, both of the terms in brackets in expression (17) are strictly increasing in  $a_1(c_H)$ . Therefore,  $\frac{\partial U_1(c_H)}{\partial a_1(c_H)}$  is everywhere strictly decreasing and for either value of  $a_2(c_H, a_1 = 1)$  there is at most one value of  $a_1(c_H)$  such that  $\frac{\partial U_1(c_H)}{\partial a_1(c_H)} = 0$ . We write:

$$\left. \frac{\partial U_1(c_H)}{\partial a_1(c_H)} \right|_{a_1(c_H)=0} = u(w_1) - q + (1-q) \left( \frac{1}{2} - \left( \frac{(1-q)^2}{q^2 + (1-q)^2} \right) \right) + qa_2(c_H, a_1 = 1) \left( u(w_2) - \frac{1}{2} \right).$$

We define:

$$\lambda(u(w_2)) = q - \frac{(1-q)(q - \frac{1}{2})}{q^2 + (1-q)^2} - qa_2(c_H, a_1 = 1) \left( u(w_2) - \frac{1}{2} \right).$$

Note that  $\lambda(u(w_2)) < q$  will always hold. However, for large  $u(w_2)$  it can be that  $\lambda(u(w_2)) < \frac{1}{2}$ . Therefore if  $\lambda(u(w_2)) \in [\frac{1}{2}, q)$  then we define  $\lambda(u(w_2)) = \lambda^*(u(w_2))$ . If  $\lambda(u(w_2)) < \frac{1}{2}$  then we define  $\lambda^*(u(w_2)) = \frac{1}{2}$ . Therefore for  $u(w_2) > \lambda^*$  it must be that  $a_1^*(c_H) \in (0, 1)$ .

Note that  $a_2^*(c_H, a_1 = 1) = 1$  if and only if:

$$u(w_2) > \frac{q^2 a_1(c_H) + q(1-q)}{(q^2 + (1-q)^2) a_1(c_H) + 2q(1-q)}$$

which implies

$$a_1(c_H) [u(w_2)(q^2 + (1-q)^2) - q^2] > q(1-q) - u(w_2)(2q(1-q)).$$

As  $u(w_2)(q^2 + (1-q)^2) - q^2$  is negative we write

$$a_1(c_H) < \frac{q(1-q) - u(w_2)(2q(1-q))}{u(w_2)(q^2 + (1-q)^2) - q^2}. \quad (18)$$

We define the right side of expression (18) as  $\hat{a}$  such that:

$$\begin{aligned} a_1(c_H) &< \hat{a} \text{ if and only if } a_2^*(c_H, a_1 = 1) = 1 \\ a_1(c_H) &\geq \hat{a} \text{ if and only if } a_2^*(c_H, a_1 = 1) = 0. \end{aligned}$$

Observe that  $\hat{a} = 0$  when  $u(w_2) = \frac{1}{2}$ ,  $\hat{a} = 1$  when  $u(w_2) = q$  and that  $\hat{a}$  is strictly increasing on this region, as  $\frac{\partial \hat{a}}{\partial u(w_2)} = \frac{q(1-q)(2q-1)}{(q^2 - u(w_2)(2q^2 - 2q + 1))^2} > 0$ .

Now consider that  $u(w_1) = u(w_2)$ . If  $a_2^*(c_H, a_1 = 1) = 1$  then

$$u(w_1) = u(w_2) \leq \frac{q - \frac{(1-q)(q - \frac{1}{2})}{q^2 + (1-q)^2} + \frac{1}{2}q}{1 + q} = \lambda$$

is equivalent to  $a_1^*(c_H) = 0$ . If  $a_2^*(c_H, a_1 = 1) = 0$  then

$$u(w_1) = u(w_2) \leq q - \frac{(1-q)(q - \frac{1}{2})}{q^2 + (1-q)^2} = \bar{\lambda}$$

is equivalent to  $a_1^*(c_H) = 0$  where  $\underline{\lambda} < \bar{\lambda}$ . By expression (16) it will then be optimal for  $a_1^*(c_H)$  to be determined where  $a_2^*(c_H, a_1 = 1) = 1$ .

■

**Lemma 6** *For every  $u(w_2) \in (q, \frac{q^2}{q^1 + (1-q)^2})$  there is a  $\mu^*(u(w_2)) \in [q, \frac{q^2}{q^1 + (1-q)^2}]$  such that for all  $u(w_1) \in (q, \mu^*(u(w_2)))$  then it is SJO that  $a_1^*(c_H) \in (0, 1)$  and  $a_1^*(c_L) = 1$  and for all  $u(w_1) \in (\mu^*(u(w_2)), \frac{q^2}{q^2 + (1-q)^2})$  then it is SJO that  $a_1^*(c_H) = a_1^*(c_L) = 1$ .*

**Proof of Lemma 6:** By Lemma 2 it must be that  $a_1^*(c_L) = 1$ . This implies that  $E[c|c_H, a_1 = 0] = \frac{q^2}{q^2 + (1-q)^2}$  therefore  $a_2^*(c_H, a_1 = 0) = 0$ . We write the utility as:

$$\begin{aligned} U_1(c_H) &= a_1(c_H)(u(w_1) - q) \\ &+ (1-q)a_1(c_H)(u(w_2) - \left( \frac{q(1-q)a_1(c_H) + (1-q)^2}{2q(1-q)a_1(c_H) + (q^2 + (1-q)^2)} \right)) \\ &\quad + (1-q)(1-a_1(c_H))(u(w_2) - \frac{1}{2}) \\ &+ qa_1(c_H)(u(w_2) - \left( \frac{q^2a_1(c_H) + q(1-q)}{(q^2 + (1-q)^2)a_1(c_H) + 2q(1-q)} \right)). \end{aligned}$$

Since this matches expression (16) in the proof of Lemma 5 with the exception that  $a_2^*(c_H, a_1 = 1) = 1$ , much of the reasoning, without the complications of determining  $a_2^*(c_H, a_1 = 1)$ , carries over. Expression (17) is also valid here:

$$\begin{aligned} \frac{\partial U_1(c_H)}{\partial a_1(c_H)} &= u(w_1) - q \\ &+ (1-q) \left( 0.5 - \left[ \frac{2q^2(1-q)^2a_1(c_H)^2 + 2q(1-q)(q^2 + (1-q)^2)a_1(c_H) + (1-q)^2(q^2 + (1-q)^2)}{4q^2(1-q)^2a_1(c_H)^2 + 4q(1-q)(q^2 + (1-q)^2)a_1(c_H) + (q^2 + (1-q)^2)^2} \right] \right) \\ &\quad + qa_1(c_H) \left( u(w_2) - \left[ \frac{q^2(q^2 + (1-q)^2)a_1(c_H)^2 + 4q^3(1-q)a_1(c_H) + 2q^2(1-q)^2}{(q^2 + (1-q)^2)^2a_1(c_H)^2 + 4q(1-q)a_1(c_H) + 4q^2(1-q)^2} \right] \right). \end{aligned}$$

Again,  $\frac{\partial U_1(c_H)}{\partial a_1(c_H)}$  is strictly decreasing and equals zero at most once. It can never be that  $a_1^*(c_H) = 0$  because

$$\left. \frac{\partial U_1(c_H)}{\partial a_1(c_H)} \right|_{a_1(c_H)=0} = (1-q) \left( 0.5 - \frac{(1-q)^2}{q^2 + (1-q)^2} \right) > 0.$$

Therefore,



$$\begin{aligned} \left. \frac{\partial U_1(c_H)}{\partial a_1(c_H)} \right|_{a_1(c_H)=1} &= u(w_1) - q + (1 - q)(0.5 + 2q^3 - 3q^2 + 2q - 1) \\ &+ qu(w_2) - \left( \frac{q^2(3 - 2q)}{8q^2(1 - q)^2 + 1} \right). \end{aligned} \quad (19)$$

And so we define:

$$\mu(u(w_2)) = \left( \frac{q + (1 - q)(0.5 - 2q^3 + 3q^2 - 2q)}{1 + q} \right) + \left( \frac{q^2(3 - 2q)}{8q^2(1 - q)^2 + 1} \right). \quad (20)$$

For  $\mu(u(w_2)) \in [q, \frac{q^2}{q^2+(1-q)^2}]$  we set  $\mu(u(w_2)) = \mu^*(u(w_2))$ . If  $\mu(u(w_2)) < q$  (because  $u(w_2)$  is in the high end of the range) then we set  $\mu^*(u(w_2)) = q$ . In this case  $a_1^*(c_H) = 1$  is *SJO* for every  $u(w_1) \in (q, \frac{q^2}{q^2+(1-q)^2})$ . If  $\mu(u(w_2)) > \frac{q^2}{q^2+(1-q)^2}$  (because  $u(w_2)$  is in the low end of the range) then we set  $\mu^*(u(w_2)) = \frac{q^2}{q^2+(1-q)^2}$ . In this case  $a_1^*(c_H) \in (0, 1)$  is *SJO* for every  $u(w_1) \in (q, \frac{q^2}{q^2+(1-q)^2})$ .

From expression (20) note that for every value of  $q \in (0.5, 1)$  it will be that

$$\frac{q^2}{q^2 + (1 - q)^2} > \mu(u(w_2)).$$

However, for some values of  $q$ ,

$$\mu(u(w_2)) > q$$

is violated.

■

**Lemma 7** *If  $u(w_1)$  and  $u(w_2) > \frac{q^2}{q^2+(1-q)^2}$  then it is *SJO* that  $a_1^*(c_H) = a_1^*(c_L) = 1$ .*

**Proof of Lemma 7:** By Lemma 2 it must be that  $a_1^*(c_L) = 1$ . This implies that  $E[c|c_H, a_1 = 0] = \frac{q^2}{q^2+(1-q)^2}$  therefore  $a_2(c_H, a_1 = 0) = 1$ . We write the utility as:

$$\begin{aligned} U_1(c_H) &= a_1(c_H)(u(w_1) - q) \\ &+ (1 - q)a_1(c_H)(u(w_2) - \left( \frac{q(1 - q)a_1(c_H) + (1 - q)^2}{2q(1 - q)a_1(c_H) + (q^2 + (1 - q)^2)} \right)) \\ &+ (1 - q)(1 - a_1(c_H))(u(w_2) - 0.5) \\ &+ qa_1(c_H)(u(w_2) - \left( \frac{q^2a_1(c_H) + q(1 - q)}{(q^2 + (1 - q)^2)a_1(c_H) + 2q(1 - q)} \right)) \\ &+ q(1 - a_1(c_H))(u(w_2) - \left( \frac{q^2}{(q^2 + (1 - q)^2)} \right)). \end{aligned}$$

Therefore

$$\begin{aligned} & \frac{\partial U_1(c_H)}{\partial a_1(c_H)} = u(w_1) - q \\ & + (1-q) \left( 0.5 - \left[ \frac{2q^2(1-q)^2 a_1(c_H)^2 + 2q(1-q)(q^2 + (1-q)^2) a_1(c_H) + (1-q)^2 (q^2 + (1-q)^2)}{4q^2(1-q)^2 a_1(c_H)^2 + 4q(1-q)(q^2 + (1-q)^2) a_1(c_H) + (q^2 + (1-q)^2)^2} \right] \right) \\ & + q \left( \frac{q^2}{q^2 + (1-q)^2} - \left[ \frac{q^2(q^2 + (1-q)^2) a_1(c_H)^2 + 4q^3(1-q) a_1(c_H) + 2q^2(1-q)^2}{(q^2 + (1-q)^2)^2 a_1(c_H)^2 + 4q(1-q) a_1(c_H) + 4q^2(1-q)^2} \right] \right). \end{aligned}$$

Again by Lemma 3 the function  $\frac{\partial U_1(c_H)}{\partial a_1(c_H)}$  is strictly decreasing. It follows from the proof of Lemma 6 that for all  $a_1(c_H)$ :

$$\frac{\partial U_1(c_H)}{\partial a_1(c_H)} > 0$$

and so it must be that  $a_1^*(c_H) = 1$ .

■

### 6.3 Proofs of Main Results

**Proof of Proposition 1:** By way of contradiction, suppose that  $a_1^*(c_H) = a_1^*(c_L) = z \in (0, 1)$ . This implies that

$$\begin{aligned} E[c|c_H, a_1 = 1] &= E[c|c_H, a_1 = 0] = E[c|c_H] = q \\ E[c|c_L, a_1 = 1] &= E[c|c_L, a_1 = 0] = E[c|c_L] = 1 - q \end{aligned}$$

and so:

$$\begin{aligned} a_2^*(c_H, a_1 = 0) &= a_2^*(c_H, a_1 = 1) = a_2^*(c_H) \\ a_2^*(c_L, a_1 = 0) &= a_2^*(c_L, a_1 = 1) = a_2^*(c_L). \end{aligned}$$

For *SJO*:

$$U_1(c_L, a_1(c_H) = z) \geq U_1(c_L, a_1(c_H) = z') \text{ for all other } z' \text{ and} \quad (21)$$

$$U_1(c_H, a_1(c_L) = z) \geq U_1(c_H, a_1(c_L) = z') \text{ for all other } z' \quad (22)$$

yielding

$$\begin{aligned} U_1^*(c_L) &= z(u(w_1) - (1-q)) + qa_2^*(c_L)[u(w_2) - (1-q)] + (1-q)a_2^*(c_H)[u(w_2) - q] \\ U_1^*(c_H) &= z(u(w_1) - q) + qa_2^*(c_H)[u(w_2) - q] + (1-q)a_2^*(c_L)[u(w_2) - (1-q)]. \end{aligned}$$

If  $u(w_1) - (1-q) > 0$  then increasing  $z$  increases both quantities in and therefore expressions (21) and (22) cannot hold. If  $u(w_1) - q < 0$  then reducing  $z$  will increase both values of  $U_1$  and therefore expressions (21) and (22) cannot hold. If  $u(w_1) - (1-q) > 0 > u(w_1) - q$  then according to Lemma 2 there exists a profitable deviation to  $a_1(c_L) = 1$  and therefore

expression (21) cannot hold. Therefore the strategy cannot be *SJO*.

■

**Proof of Proposition 5:** With a perfect memory agent the second period posteriors and expectations are standard:

$$\begin{aligned} E[c|s_1 = c_H, s_2 = c_H] &= \frac{q^2}{q^2 + (1-q)^2} \\ E[c|s_1 = c_H, s_2 = c_L] &= \frac{1}{2} \\ E[c|s_1 = c_L, s_2 = c_H] &= \frac{1}{2} \\ E[c|s_1 = c_L, s_2 = c_L] &= \frac{(1-q)^2}{q^2 + (1-q)^2} \end{aligned}$$

as first period actions do not affect second period beliefs. The ex-ante utility of the agent is:

$$\begin{aligned} U_0(w) &= \frac{1}{2} \{a_1(c_L)[u(w_1) - (1-q)] \\ &+ qa_2(c_L, c_L)[u(w_2) - \frac{(1-q)^2}{q^2 + (1-q)^2}] + (1-q)a_2(c_L, c_H)[u(w_2) - \frac{1}{2}]\} \\ &+ \frac{1}{2} \{a_1(c_H)[u(w_1) - q] \\ &+ qa_2(c_H, c_H)[u(w_2) - \frac{q^2}{q^2 + (1-q)^2}] + (1-q)a_2(c_H, c_L)[u(w_2) - \frac{1}{2}]\}. \end{aligned}$$

From this it follows that the  $a_1^*(c_L)$  and  $a_1^*(c_H)$  are selected as if the agent was myopic. Suppose that Definitions 2 and 3 are satisfied. Together with the observation of myopic behavior, closeness implies that  $a_1^{*C}(c_H) = a_1^{*I}(c_H)$  and  $a_1^{*C}(c_L) = a_1^{*I}(c_L)$ . Also from closeness it must be that:

$$a_2^{*C}(s_1, s_2) = a_2^{*I}(s_1, s_2) \text{ for } s_1, s_2 \in \{c_L, c_H\}.$$

Therefore, expression (4) in Definition 2:

$$\begin{aligned} &u(w_1^C)\Psi_1^C + u(w_2^C)\Psi_2^C \\ &> u(w_1^I)\Psi_1^I + u(w_2^I)\Psi_2^I \end{aligned}$$

is equivalent to

$$U_0(w^C) - U_0(w^I) > 0$$

and the agent cannot display *PIP*.

■

**Proof of Proposition 6:** Because first period and second period behavior are identical

in both contracts we can write:

$$\begin{aligned}
U_0(w^C) - U_0(w^I) &= \left( \frac{a_1(c_L) + a_1(c_H)}{2} \right) (u(w_1^C) - u(w_1^I)) \\
&\quad + \frac{1}{2} \{ (qa_1(c_H) + (1-q)a_1(c_L))a_2(c_H, a_1 = 1) \\
&\quad + (q(1 - a_1(c_H)) + (1-q)(1 - a_1(c_L)))a_2(c_H, a_1 = 0) \\
&\quad + ((1-q)a_1(c_H) + qa_1(c_L))a_2(c_L, a_1 = 1) \\
&\quad + ((1-q)(1 - a_1(c_H)) + q(1 - a_1(c_L)))a_2(c_L, a_1 = 0) \} (u(w_2^C) - u(w_2^I)).
\end{aligned}$$

This can be rewritten as expression (4) in Definition 2 and so it must be that

$$U_0(w^C) - U_0(w^I) > 0$$

and so *PIP* can never occur.

■

**Proof of Proposition 2:** If  $u(w_1), u(w_2) > \frac{q^2}{q^1 + (1-q)^2}$  then Lemma 7 shows that  $a_1^*(c_L) = a_1^*(c_H) = 1$ . If  $u(w_1), u(w_2) \in (1-q, \frac{1}{2}]$  then Lemma 4 shows that  $a_1^*(c_L) = 1$  and  $a_1^*(c_H) = 0$ . Under both cases Proposition 6 applies and so Proposition 2 is proved.

■

**Proof of Proposition 3:** Lemma 5 shows that such a  $\lambda^*$  exists. Further Lemma 5 shows that for  $u(w_1), u(w_2) \in (\frac{1}{2}, \lambda^*)$  that  $a_1^*(c_L) = 1$  and  $a_1^*(c_H) = 0$ . Lemma 6 shows that such a  $\mu^*$  exists and that for  $u(w_1), u(w_2) \in (\mu^*, \frac{q^2}{q^1 + (1-q)^2})$  that  $a_1^*(c_L) = a_1^*(c_H) = 1$ . Under both cases Proposition 6 applies and so Proposition 3 is proved.

■

The following Lemmas will be helpful in the Proof of Proposition 4.

**Lemma 8** *Suppose that the SJO is such that  $a_1^*(c_L) = 1$  and  $a_1^*(c_H) \in (0, 1)$  for  $w$  and  $w'$ . If*

$$u(w_1) + qu(w_2) > u(w_1') + qu(w_2')$$

*then  $a_1^*(c_H) > a_1'^*(c_H)$*

**Proof:** By assumption,  $a_1^*(c_H)$  has an interior maximum which is determined by:

$$\left. \frac{\partial U_1(c_H)}{a_1(c_H)} \right|_{a_1^*(c_H)} = 0$$

where

$$\begin{aligned} & \frac{\partial U_1(c_H)}{a_1(c_H)} = u(w_1) - q \\ & + (1-q) \left( 0.5 - \left[ \frac{2q^2(1-q)^2 a_1(c_H)^2 + (q^2 + (1-q)^2)(2q(1-q)a_1(c_H) + (1-q)^2)}{(2q(1-q)a_1(c_H) + q^2 + (1-q)^2)^2} \right] \right) \\ & + q \left( u(w_2) - \left[ \frac{q^2(q^2 + (1-q)^2) a_1(c_H)^2 + (2q(1-q))(2q^2 a_1(c_H) + q(1-q))}{((q^2 + (1-q)^2)a_1(c_H) + 2q(1-q))^2} \right] \right). \end{aligned}$$

By Lemma 3 the expressions in the brackets are strictly increasing on  $a_1(c_H) \in [0, 1]$  for every  $q \in (\frac{1}{2}, 1)$ . Since

$$u(w_1) - u(w'_1) + q(u(w_2) - u(w'_2)) > 0$$

it follows that  $a_1^*(c_H) > a_1'^*(c_H)$ .

■

**Lemma 9** *If  $a_1^*(c_L) = 1$  and  $a_1^*(c_H) \in (0, 1)$  then for every  $\hat{a}_1(c_H)$  it will be that*

$$\left. \frac{\partial U_1(c_H)}{\partial a_1(c_H)} \right|_{\hat{a}_1(c_H)} - \left. \frac{\partial U_0}{\partial a_1(c_H)} \right|_{\hat{a}_1(c_H)} = q(1-q)(2q-1) > 0.$$

**Proof:** For these parameter values we can write:

$$\begin{aligned} & \frac{\partial U_1(c_H)}{a_1(c_H)} = u(w_1) - q \\ & + (1-q) \left( 0.5 - \left[ \frac{2q^2(1-q)^2 a_1(c_H)^2 + 2q(1-q)(q^2 + (1-q)^2) a_1(c_H) + (1-q)^2(q^2 + (1-q)^2)}{4q^2(1-q)^2 a_1(c_H)^2 + 4q(1-q)(q^2 + (1-q)^2) a_1(c_H) + (q^2 + (1-q)^2)^2} \right] \right) \\ & + q \left( u(w_2) - \left[ \frac{q^2(q^2 + (1-q)^2) a_1(c_H)^2 + 4q^3(1-q) a_1(c_H) + 2q^2(1-q)^2}{(q^2 + (1-q)^2)^2 + 4q(1-q)(q^2 + (1-q)^2) a_1(c_H) + 4q^2(1-q)^2} \right] \right) \end{aligned}$$

and

$$\begin{aligned} & \frac{\partial U_0}{\partial a_1(c_H)} = u(w_1) - q \\ & + (1-q) \left( 0.5 - \left( \frac{2q^2(1-q)^2 a_1(c_H)^2 + 2q(1-q)(q^2 + (1-q)^2) a_1(c_H) + q^4 + (1-q)^4}{4q^2(1-q)^2 a_1(c_H)^2 + 4q(1-q)(q^2 + (1-q)^2) a_1(c_H) + (q^2 + (1-q)^2)^2} \right) \right) \\ & + q \left( u(w_2) - \left( \frac{q^2(q^2 + (1-q)^2) a_1(c_H)^2 + 4q^3(1-q) a_1(c_H) + (1-q)^2(2q^2 + 2q - 1)}{(q^2 + (1-q)^2)^2 + 4q(1-q)(q^2 + (1-q)^2) a_1(c_H) + 4q^2(1-q)^2} \right) \right). \end{aligned}$$

Mercifully the difference simplifies to:

$$\frac{\partial U_1(c_H)}{\partial a_1(c_H)} - \frac{\partial U_0}{\partial a_1(c_H)} = (1-q)(-q^2 + 2q^3) + q(1-q)^2(2q-1).$$

So the Lemma is proved.

■

**Proof of Proposition 4:** We offer a constructive proof where we identify the process of determining a contract  $w^I$  sufficient to satisfy *PIP*. We proceed in two main parts. Part 1 shows that a contract  $w^I$  which induces first period actions so that  $a_1^C > a_1^I$  will imply that

$$2(U_0(w^C) - U_0(w^I)) - \varepsilon < 0.$$

Part 2 shows that a contract  $w^I$  can be found which makes  $\varepsilon$  arbitrarily small.

**Part 1:**

For the proof to follow it must be that  $a_1^*(c_H) \in (0, 1)$  and  $a_1^*(c_L) = 1$ . This is the content of Lemmas 5 and 6. In the exhibition of *PIP* it is required that given a constant contract  $w^C$  we can find an increasing contract  $w^I$  preferred by the ex-ante player but satisfying the requirements provided in the definition. Given the increasing contract  $w^C$ , we select the increasing contract  $w^I$  such that the former is more valuable:

$$\begin{aligned} & \Psi_1^C u(w_1^C) + \Psi_2^C u(w_2^C) \\ &= \Psi_1^I u(w_1^I) + \Psi_2^I u(w_2^I) + \varepsilon. \end{aligned}$$

We can then rewrite the difference in ex-ante utilities as

$$\begin{aligned} & 2(U_0(w^C) - U_0(w^I)) - \varepsilon \tag{23} \\ &= a_1^I(c_H)q - a_1^C(c_H)q \\ &+ ((1-q)a_1^I(c_H) + q) \left( \frac{q(1-q)a_1^I(c_H) + (1-q)^2}{2q(1-q)a_1^I(c_H) + (q^2 + (1-q)^2)} \right) \\ &- (1-q)a_1^C(c_H) + q) \left( \frac{q(1-q)a_1^C(c_H) + (1-q)^2}{2q(1-q)a_1^C(c_H) + (q^2 + (1-q)^2)} \right) \\ &+ (1-q) \frac{1}{2} (a_1^C(c_H) - a_1^I(c_H)) \\ &+ (qa_1^I(c_H) + (1-q)) \left( \frac{q^2 a_1^I(c_H) + q(1-q)}{(q^2 + (1-q)^2)a_1^I(c_H) + 2q(1-q)} \right) \\ &- (qa_1^C(c_H) + (1-q)) \left( \frac{q^2 a_1^C(c_H) + q(1-q)}{(q^2 + (1-q)^2)a_1^C(c_H) + 2q(1-q)} \right). \end{aligned}$$

To see that  $\frac{\partial U_0}{\partial a_1(c_H)}$  is monotonically decreasing, note (as we have established) that  $\frac{\partial U_1}{\partial a_1(c_H)}$  is monotonically decreasing. By Lemma 9  $\frac{\partial U_1}{\partial a_1(c_H)} - \frac{\partial U_0}{\partial a_1(c_H)}$  equals a positive constant. Therefore the right side of (23) is negative for any  $a_1^C(c_H) > a_1^I(c_H)$  and so the balance of the proof consists in finding an  $w^I$  which induces an appropriately small  $\varepsilon$ .

**Part 2:**

We have a limited amount of room in which to find  $w^I$ . For (i) the contract  $w^I$  must be such that  $u(w_1^I), u(w_2^I) \in (\lambda^*, q)$ . For (ii) the contract  $w^I$  it must be that  $q < \mu^*$  and  $u(w_1^I), u(w_2^I) \in (q, \mu^*)$ . For convenience, we denote the lower element of these sets as  $\underline{u}$  and the upper element as  $\bar{u}$ .

Pick an preliminary  $u(w_1^I)^0 = u(w_2^I)^0 = u(w^I)^0$  such that

$$u(w^I)^0 < u(w^C)$$

and

$$\begin{aligned} & (\Psi_1^C + \Psi_2^C)u(w_2^C) \\ = & (\Psi_1^{I0} + \Psi_2^{I0})u(w^I)^0 + \delta. \end{aligned}$$

We may pick  $\delta$  as small as we like because  $(\Psi_1^{I0} + \Psi_2^{I0})u(w^I)^0$  is an increasing function of  $u(w^I)^0$ . Furthermore, by Lemma 8 any such  $u(w^I)^0$  induces  $a_1^C(c_H) > a_1^I(c_H)$ . We can select a new  $u(w_1^I)$  and  $u(w_2^I)$  without affecting  $a_1^I(c_H)$  (therefore not affecting the right hand side of expression (23))

$$u(w_1^I) + qu(w_2^I) = (1 + q)u(w^I)^0.$$

Define

$$f(x) = \Psi_1^I(u(w^I)^0 - qx) + \Psi_2^I(u(w_2^I)^0 + x)$$

where  $x \geq 0$ . As noted above we have a limited amount of room with which to work. In order to stay in the allowed region it must be that  $x < x^*$  where

$$x^* = \min(\bar{u} - u(w^I)^0, q(u(w^I)^0 - \underline{u})).$$

So we require that  $x \in [0, x^*)$ . Also note that:

$$\frac{\partial f}{\partial x} = \frac{1}{2} + q + \frac{a_1^I(c_H)}{2}.$$

So for every  $a_1^I(c_H)$ , this value is at least  $\frac{1}{2} + q$ . Given  $x^*$  and  $q$  we can pick  $u(w^I)^0$  such that

$$\left(\frac{1}{2} + q\right)x^* > \delta.$$

This implies that

$$f(x^*) > \Psi_1^C u(w_1^C) + \Psi_2^C u(w_2^C)$$

for  $x^*$ . Therefore there are values of  $x \in (0, x^*)$  such that  $\varepsilon$  in expression (23) can be made arbitrarily small. Therefore the Proposition is proved.

■

## 7 References

Arkes, Hal and Harkness, Allan (1980): "Effect of Making a Diagnosis on Subsequent Recognition of Symptoms," *Journal of Experimental Psychology: Human Learning and Memory*, 6(5): 568-575.

Bem, Daryl (1972): "Self-Perception Theory," in *Advances in Experimental Social Psychology*, Berkowitz, L. (Ed.), New York, Academic Press, 1-62.

Benabou, Roland and Tirole Jean (2003): "Intrinsic and Extrinsic Motivation," *Review of Economic Studies*, 70, 489-520.

Benabou, Roland and Tirole, Jean (2004): "Willpower and Personal Rules," *Journal of Political Economy*, 112 (4), 848-886.

Bernheim, Douglas and Thomadsen, Raphael (2005): "Memory and Anticipation," *Economic Journal*, 115, 271-304.

Bradley, Darren (2003): "Sleeping Beauty: a Note on Dorr's Argument for 1/3," *Analysis*, 63(3), 266-268.

Clark, Andrew (1999): "Are Wages Habit-Forming? Evidence from Micro Data," *Journal of Economic Behavior and Organization*, 39: 179-200.

Dorr, Cian (2002): "Sleeping Beauty: in Defense of Elga," *Analysis*, 62(4), 292-296.

Dow, James (1991): "Search Decisions with Limited Memory," *Review of Economic Studies*, 58: 1-14.

Elga, Adam (2000): "Self-locating Belief and the Sleeping Beauty Problem," *Analysis*, 60(2), 143-147.

Eyster, Erik (2002): "Rationalizing the Past: A Taste for Consistency," unpublished Oxford University.

Gabrielcik, Adele and Fazio, Russell (1984): "Priming and Frequency Estimation: A Strict Test of the Availability Heuristic," *Personality and Social Psychology Bulletin*, 10(1): 85-89.

Gigliotti, Gary and Sopher, Barry (1997): "Violations in Present-Value Maximization in Income Choice," *Theory and Decision*, 43, 45-69.

Gigliotti, Gary and Sopher, Barry (2003): "Analysis of Intertemporal Choice: a New Framework and Experimental Results," *Theory and Decision*, 55: 209-233.



Grund, Christian and Sliwka, Dirk (2007): "Reference-Dependent Preferences and the Impact of Wage Increases on Job Satisfaction: Theory and Evidence," *Journal of Institutional and Theoretical Economics*, 163: 313-335.

Guysé, Jeffrey, Keller, L. Robin and Eppel, Thomas (2002): "Valuing Environmental Outcomes: Preferences for Constant or Improving Sequences," *Organizational Behavior and Human Decision Processes*, 87(2): 253-277.

Hirshleifer, David and Welch, Ivo (2002): "An Economic Approach to the Psychology of Change: Amnesia, Inertia and Impulsiveness," *Journal of Economics and Management Strategy*, 11(3), 379-421.

Lee, Dong and Uhlemann, Max (1994): "Development of an Instrument to Assess Clinicians' Recognition Memory Using a Videotaped Stimulus Interview," *Journal of Clinical Psychology*, 50(5): 802-809.

Lewis, David (2001): "Sleeping Beauty: Reply to Elga," *Analysis*, 61(3), 171-176.

Loewenstein, George and Sicherman, Nachum (1991): "Do Workers Prefer Increasing Wage Profiles?" *Journal of Labor Economics*, 9, 67-84.

Manzini, Paola, Mariotti, Marco and Mittone, Luigi (2006): "Choosing Monetary Sequences: Theory and Experimental Evidence," unpublished University of London and University of Trento.

Matsumoto, Dawn, Peecher, Mark and Rich, Jay (2000): "Evaluations of Outcome Sequences," *Organizational Behavior and Human Decision Processes*, 83(2), 331-352.

Monton, Bradley (2002): "Sleeping Beauty and the Forgetful Bayesian," *Analysis*, 62(1), 47-53.

Mullainathan, Sendhil (2002): "A Memory-Based Model of Bounded Rationality," *Quarterly Journal of Economics*, 117 (3), 735-774.

Piccione, Michele and Rubenstein, Ariel (1997): "On the Interpretation of Decision Problems with Imperfect Recall," *Games and Economic Behavior*, 20, 3-24.

Read, Daniel and Powell, Melanie (2002): "Reasons for Sequence Preferences," *Journal of Behavioral Decision Making*, 15: 433-460.

Schwarz, Norbert, Bless, Herbert, Strack, Fritz, Klumpp, Gisela and Rittenauer-Schatka, Helga (1991): "Ease of Retrieval as Information: Another Look at the Availability Heuristic," *Journal of Personality and Social Psychology*, 61(2): 195-202.

Swank, Otto (2006): "The Self-Perception Theory Versus a Dynamic Learning Model," unpublished Erasmus University Rotterdam.

Tversky, Amos and Kahneman, Daniel (1973): "Availability: A Heuristic for Judging Frequency and Probability," *Cognitive Psychology*, 5: 207-232.

Weintraub, Ruth (2004): "Sleeping Beauty: a Simple Solution," *Analysis*, 64(4), 8-10.

Wilson, Andrea (2004): "Bounded Memory and Biases in Information Processes," unpublished Harvard.

Yariv, Leeat (2006): "I'll See It When I Believe It- A Simple Model of Cognitive Consistency," unpublished California Institute of Technology.