An Axiomatic Model of Cognitive Dissonance

YUZHAO YANG
Department of Economics, Boston University

CLICK HERE FOR THE LATEST VERSION.

Cognitive dissonance is the internal tension individuals feel when their past actions are inconsistent with their beliefs or attitudes. I propose an intrapersonal game to model a decision-maker who distorts her beliefs to mitigate cognitive dissonance from past choices. Two selves make sequential, observable decisions, with unobservable belief manipulation occurring in the interim stage between them. The subgame perfect Nash equilibria are characterized by tractable axioms on choice patterns, with parameters identifiable from choice data. I demonstrate that the model is useful for studying path dependence in decision-making. The model matches a variety of experimental and real-world evidence consistent with cognitive dissonance theory. Applying the model provides new insights into existing topics, including add-on selling, behavioral poverty traps, and the value of information.

KEYWORDS. Cognitive Dissonance. Intrapersonal Equilibrium. Motivated Reasoning.

Yuzhao Yang: allenyyz@bu.edu
I am grateful to Barton Lipman, Jawwad Noor, and Larry Epstein for their guidance and support and to Marciano Siniscalchi for many insightful discussions. I also thank Nina Bobkova, Xiaoyu Cheng, Fernando Payro Chew, Paul W. Dai, Yifan Dai, Krishna Dasaratha, Eddie Dekel, Alexander Jakobsen, Asen Kochov, Peter Klibanoff, Albert Ma, Andrew Newman, Juan Ortner, Enrico Salonia, Tomasz Strzalecki, Gerelt Tserenjigmid, and seminar participants at Oxford, the 3B Theory Workshop, D-TEA Paris, and the Stony Brook Game Theory Conference for their valuable feedback and discussions.

He that has once done you a kindness
will be more ready to do you another...

Benjamin Franklin

1. Introduction

Past actions can influence current views and affect decisions for the future. For example, individuals initially choosing high-risk industries may gradually become desensitized to dangers over time, caring less about safety regulations (Akerlof and Dickens, 1982). Similarly, agreeing to a small favor can increase the likelihood of complying with larger requests later, a concept known as the *foot-in-the-door phenomenon* (Gneezy et al., 2012). After incurring a loss on an investment, investors may hesitate to terminate it, struggling to accept the past decision as a mistake – an example of *effort justification* (Staw, 1981, Chang et al., 2016).

The psychological theory of *cognitive dissonance* explains the scenarios described above. In each case, the initial decision is made under uncertainty, which can give rise to lingering doubts about potential adverse outcomes: "I might have chosen a job that could result in serious injury," "I might have done a favor for the wrong person," or "I might be an incompetent investor holding worthless assets." These doubts create internal tensions known as cognitive dissonance (Festinger, 1957). After the initial decision, the need to reduce cognitive dissonance shapes the decision-maker's (DM's) beliefs and subsequent choices.

I introduce the cognitive dissonance (CD) game, an intrapersonal game in which three forward-looking players sequentially choose an action, select a belief, and then take a subsequent action. I make three contributions. First, through introducing a model that applies to general two-period decision problems, I provide a useful analytical tool for exploring new implications of cognitive dissonance, including add-on selling (Ellison, 2005, Gabaix and Laibson, 2006), cognitive poverty traps (Bertrand et al., 2004, Mani et al., 2013), and information avoidance (Golman et al., 2017) as discussed in this paper. Second, besides generating new implications, my model generates various stylized facts in the cognitive dissonance literature. Third, I provide axioms on observable choices that characterize the model and show that the parameters can be identified from choice data.

The CD game unfolds as follows. The true state is drawn from the state space Ω and is fixed throughout the game. Three players move sequentially: Player 1, 2, and an intermediate player who moves between them. Player 1 has belief μ_1 over the state space Ω. She makes the initial move, determining the state-dependent consumption $c_1^* = (c_{1\omega}^*)_{\omega\in\Omega}$ and the constraint C_2 for future consumption. Let (c_1^*, C_2) represent this initial decision. Player 1 correctly foresees the equilibrium choice $c_2^* \in C_2$ made by future self, with payoff given by the time-separable subjective expected utility (SEU)

$$U_{\mu_1}(c_1^*, c_2^*) = \sum_{t\in\{1,2\}} \delta^{t-1}\mathbb{E}_{\mu_1}[u(c_t^*)]. \tag{1}$$

After Player 1 selects (c_1^*, C_2), an intermediate player optimally chooses a belief to enhance her perception of the previously chosen action. Correctly foreseeing the equilibrium period-2 choice c_2^*, her payoff from choosing belief $\mu \in \Delta\Omega$ is given by

$$\theta U_\mu(c_1^*, c_2^*) - D_{KL}(\mu||\mu_1). \tag{2}$$

Objective functions similar to (2) have been used to study both pessimism (Hansen and Sargent, 2001, Strzalecki, 2011) and optimism (Caplin and Leahy, 2019) in a static setting. I employ this formulation in the dynamic setting to capture the desire to mitigate the dissonance after performing the action (c_1^*, C_2). The term $\theta U_\mu(c_1^*, c_2^*)$ measures the value of the past decision (c_1^*, C_2). The intermediate player maximizes this value by downplaying the likelihood of undesirable outcomes from the past decision, reflecting the desire to mitigate post-decision dissonance. I include the future consumption c_2^* to reflect the continuation value of the past decision. The intermediate player understands how her belief choice will influence the choice for c_2^* in the future.

In the objective function (2), the cost term $D_{KL}(\mu||\mu_1)$ is the Kullback-Leibler (KL) divergence from the *ex-ante* belief to the distorted belief, measuring the "distance" between them. It captures the agent's distaste for changes of opinion in the absence of information, which can cause feelings of inconsistency with past selves. The parameter $\theta \geq 0$ is referred to as the *dissonance factor*, which quantifies the desire to reduce cognitive dissonance, with the model aligning with the rational benchmark when $\theta = 0$.

Finally, Player 2 has the belief chosen by the intermediate player, which I denote as μ_2. She chooses the optimal consumption $c_2 \in C_2$ that maximizes the expected utility

$$\mathbb{E}_{\mu_2}[u(c_2)]. \tag{3}$$

The CD game has one more parameter θ than the rational benchmark. This single additional degree of freedom is sufficient to capture various established findings. For instance, as illustrated at the start of the paper, *effort justification* suggests that after investing heavily in a project, people justify their effort by convincing themselves that the task is valuable (Harmon-Jones and Mills, 2019). *Induced compliance* states that after being pressured into doing an action, individuals adjust their opinions to align with the action to reduce cognitive dissonance (Festinger and Carlsmith, 1959, Vaidis et al., 2024). Since voluntarily making a poor decision signals a greater lack of competence and undermines self-image, cognitive dissonance can be more intense when choices are made freely rather than assigned by others (Chang et al., 2016, Harmon-Jones and Mills, 2019). In Section 4, I present applications of the CD game that accounts for the empirical regularities described above.

The CD game involves multiple interacting components, but its subgame perfect Nash equilibria (SPNE) can be characterized through six tractable axioms derived from the DM's observable behaviors. Axioms 1-3 and 5-6 capture features of the standard time-separable SEU for a forward-looking agent. Axiom 4, the central axiom, allows a specific form of *dynamic inconsistency* in decision-making that is consistent with the mitigation of cognitive dissonance. My main theorem shows the DM's observable choices satisfy Axioms 1-6 *if and only if* the choices are consistent with the SPNE of the CD game. I also establish the uniqueness of my model's parameters and introduce a comparative notion for the strength of cognitive dissonance.

I discuss the following example as a concrete illustration of Axiom 4, highlighting the key choice implication of my model and distinguishing it from other forms of time-inconsistent behavior discussed in the literature.

EXAMPLE 1. A forward-looking and impatient worker is deciding whether to work in a chemical plant for $t = 1, 2$. Workers in the plant are regularly exposed to a chemical that they believe may be safe (state G) or poisonous (state B) with equal probability. The worker's decision in each period is represented by a pair (c_G, c_B) specifying the payoffs (in utils) in each state. For each period $t = 1, 2$, he may choose to work in the plant $c_t = (2, -4)$ or an outside option $\emptyset_t$. Suppose $\emptyset_1 = (-4, -4)$, reflecting an initial economic downturn, and $\emptyset_2 = (0, 0)$, reflecting a return to normal conditions. It is easily verified that the worker would choose c_1 for time $t = 1$ because the economic downturn compels him to work in the plant despite his safety concerns.

First, let us ask what the worker, at time $t = 1$, would want his time $t = 2$ self to choose. The answer is he would prefer his future self to choose $\emptyset_2$, *i.e.* to quit the job as the economic condition returns to normal.[1] However, his future self may actually choose to continue working in the plant (c_2) instead of quitting. To see why, note that the worker experiences an unpleasant feeling of fear when he works at the plant at $t = 1$ and comes into contact with the chemical. To overcome this fear, he would be motivated to distort his belief to understate the likelihood of state B ("poisonous"). At $t = 2$, if the worker's belief in B drops below $1/3$, he would choose c_2 in the second period, as its expected payoff would then exceed $\emptyset_2$.[2]

This example represents a form of dynamic inconsistency consistent with mitigating cognitive dissonance. Unlike the internalization of anticipated emotions in Caplin and Leahy (2001), retrospective dissonance emerges after making choices. It also diverges from regret minimization (Eyster et al., 2022), as the worker's decision to remain at the plant statewise dominates the outside option at $t = 1$, precluding the experience of regret across all possible states. □

I explore three applications of the CD game. The first application models *upselling*, in which a monopolist sells a primary product and then offers add-ons following the initial purchase. While cognitive dissonance allows the monopolist to charge a higher price for the add-on, its effect on the equilibrium quantity is less clear. I show that the seller optimally employs a *mass marketing* strategy, expanding the customer base for both the basic product and the add-on. Consumer welfare always strictly *exceeds* the benchmark scenario where both products are sold simultaneously – in contrast to the welfare implications of existing models of add-on selling (Ellison, 2005, Gabaix and Laibson, 2006). If the consumers' dissonance factor, θ, is sufficiently small, the upselling strategy is strictly profitable for the seller.

My second application offers a concrete and testable mechanism for the idea that poverty impairs decision-making quality by creating mental discomfort (Kremer et al., 2019). I demonstrate that, with a fixed dissonance factor θ, lower-income individuals may experience greater cognitive dissonance because they are more vulnerable to the potential negative outcomes of past choices. This heightened dissonance results in greater belief distortion, reducing decision-making quality and effectively imposing a "cognitive tax" that further deepens the challenges of poverty.

[1] That is to say, the worker would like to commit to quitting the job for his future self, if he had the power to make such a commitment.

[2] This example aligns with Akerlof and Dickens (1982), who first introduced cognitive dissonance to economics through a labor market model, reflecting real-world evidence that workers in hazardous industries systematically "underestimate hazards," "take unnecessary risks," and "lack sufficient safety awareness" (Dodoo and Al-Samarraie, 2021).

My third application extends the CD game by introducing belief updating during the interim period and examines how cognitive dissonance affects the *ex-ante* value of information. While it is commonly understood that dynamically inconsistent agents may use information avoidance as a commitment device (Carrillo and Mariotti, 2000, Golman et al., 2017), I show that cognitive dissonance can either increase or decrease the value of information. Low-quality information can elicit dissonance and harm the DM, while high-quality information reduces the dissonance associated with payoff uncertainties, thereby improving the quality of decision-making.

The rest of this paper is organized as follows. Section 2 introduces the CD game. Section 3 introduces my axiomatic system and the representation theorem. Section 4 matches my model with various empirical regularities. Section 5 discusses the three applications of my model. Section 6 discusses the rationale behind key assumptions of the CD game.

Related Literature

Cognitive Dissonance Various applied theory frameworks have been used to study cognitive dissonance in economic contexts, including labor market (Akerlof and Dickens, 1982), social values (Rabin, 1994, Konow, 2000), identity investment (Bénabou and Tirole, 2011), economic mobility (Haagsma and Koning, 2002, Oxoby, 2003, 2004, Penn, 2017), financial markets (Chang et al., 2016), political beliefs (Acharya et al., 2018), and effort justification (Suzuki, 2019). In a multi-period setting, Yariv (2005) developed a dynamic choice model in which the agent reduces dissonance arising from holding beliefs that contradict past beliefs. Compared to existing literature, this paper provides an axiomatic treatment of cognitive dissonance. My model applies to general two-period decision problems, offering broader applicability than previous models focusing on specific contexts. This abstraction enables the model to explore new applications of cognitive dissonance and account for various empirical patterns observed in classic psychological experiments, empirical studies, and recent lab experiments and replications.[3]

[3] For classic psychological experiments, see Brehm (1956), Festinger and Carlsmith (1959), Aronson and Mills (1959), Knox and Inkster (1968), Cooper and Fazio (1984), Beauvois et al. (1995); for empirical studies, see (Mullainathan and Washington, 2009, Gneezy et al., 2012, Chang et al., 2016, Orhun et al., 2024); for recent lab experiments and replications, see (Van Veen et al., 2009, Vaidis et al., 2024, Fan, 2024).

Motivated Reasoning The cognitive dissonance literature is part of the broader body of research on motivated reasoning. A strand of this literature studies motivated reasoning models with axiomatic foundations, typically in static contexts or involving myopic agents. Kovach (2020) and Mayraz (2019) characterize wishful thinking models in which SEU maximizers distort beliefs to view an *exogenous* reference point in a positive light. Bracha and Brown (2012), Caplin and Leahy (2019), and Burgh and Melo (2024) study non-SEU models in which belief distortion favors an *endogenous* optimal choice, using static analogs of the interim optimization problem (2) with different cost functions.[4] In contrast to this strand of literature, I develop a two-period multi-selves model featuring forward-looking agents. The dynamic nature of the model enables analysis of *path dependence* in decision-making, as well as commitment preferences, information value, and intertemporal trade-offs. Brunnermeier and Parker (2005) explored dynamically consistent agents selecting optimal prior beliefs, while this paper addresses dynamically inconsistent agents motivated by retrospective dissonance reduction. For related work on belief (or more broadly, preference) as an object of choice, see also Loewenstein (1987), Caplin and Leahy (2001), Köszegi (2006), Bernheim et al. (2021), Eyster et al. (2022), Chambers et al. (2023), and Sinander (2023). As stated above, the key distinction in my model is its focus on retrospective motivation in belief choice and path dependence in decision-making.[5]

Consistent Planning In the consistent planning literature, a "current self" anticipates and responds to the choices of "future selves" through backward induction (Strotz, 1955, Peleg and Yaari, 1973, Laibson, 1997, Carrillo and Mariotti, 2000, Siniscalchi, 2011, Dekel and Lipman, 2012, Jakobsen, 2021). I extend the consistent planning framework by *endogenizing* the future self's preferences to incorporate cognitive dissonance mitigation and provide the corresponding axiomatic foundation. My approach builds on Akerlof and Dickens (1982)'s pioneering work on cognitive dissonance, where choices unfold in a sequence of action, belief, and subsequent action.

Relation to Other Models Other theories in the literature emphasize the link between past actions and future beliefs or, more broadly, preferences. Eyster et al. (2022) investigates an

[4] Caplin and Leahy (2019) used the KL-divergence cost, Burgh and Melo (2024) used a ϕ-divergence cost, and Bracha and Brown (2012) considered a general convex cost function. Bracha and Brown (2012) presents a formal axiomatic characterization, shifting the quasi-concavity condition in Maccheroni et al. (2006) to quasi-convexity.

[5] A notable exception is Eyster et al. (2022), which develops a theory of *ex-post* rationalization. I will provide a detailed comparison later in this section.

agent who optimally chooses preferences to rationalize past decisions that, in hindsight, are recognized as mistakes. The key difference is that, in my model, dynamic inconsistency may arise without new information indicating past choices were mistakes. This is consistent with the mitigation of cognitive dissonance, where performing an action changes belief "while no new information is imparted" (Akerlof and Dickens, 1982). Bénabou and Tirole (2004, 2006, 2011) introduced models where Bayesian agents with limited memories learn from their past actions, which also accounts for changes in beliefs without external information arriving. The key distinction of my paper is its close alignment with the core predictions of cognitive dissonance theory, as discussed in Section 4. For example, the CD game explains induced compliance experiments (Festinger and Carlsmith, 1959, Vaidis et al., 2024), where most participants comply with the experimenter's requests and subsequently adjust their opinions to align with their actions. My model aligns with this prediction, whereas a Bayesian agent with limited memory would gain minimal information from the initial act of compliance. Decision theorists also study how agents may become increasingly pessimistic over time due to fear of executing an action ("cold feet") (Epstein and Kopylov, 2007) or concerns about model misspecification (Lanzani, 2022). However, the contribution of this paper is not limited to contrasting optimism with pessimism. As discussed in Section 4 and 5, my model captures various empirical patterns and real-life applications that cannot be explained merely by shifting between optimism and pessimism in existing models.

2. The Cognitive Dissonance Game

Basic Framework Let Ω be a finite set of states with generic elements ω and ψ, $\mathcal{E}$ the collection of all non-empty proper subsets of Ω, and $\Delta\Omega$ the set of probability distributions over Ω with generic element $\mu = (\mu_\omega)_{\omega\in\Omega}$. The payoff set X is an abstract metric space with elements x and y. For each of the two periods $t \in \mathcal{T} \equiv \{1, 2\}$, a period-$t$ consumption is a function $c_t : X \to \Omega$, which is also written as $(c_{t\omega})_{\omega\in\Omega}$.[6] Let $\mathcal{C}_t$ be the set of all period-t consumptions. A (state-dependent) consumption stream is $c = (c_1, c_2) \in \mathcal{C}_1 \times \mathcal{C}_2 \equiv \mathcal{C}$.

Action and Menu A period-1 action is a pair $a = (c_1, C_2)$, where $c_1 \in \mathcal{C}_1$ represents period-1 consumption and $C_2 \subset \mathcal{C}_2$ is a non-empty, finite set of period-2 consumptions, referred to

[6] Although c_t is commonly referred to as an *act* in the literature, this paper will consistently use *consumption* to avoid confusion with the "period-1 action" defined later. I also use $x \in X$ to denote consumption that pays x in every state.

as the period-2 menu. The set of all period-1 menus is denoted by $\mathcal{A}$.[7] To simplify notation, I use $c = (c_1, c_2) \in \mathcal{C}$ to represent the period-1 action $(c_1, \{c_2\})$. Let $\mathbb{A}$ denote the set of all period-1 actions. A period-1 menu A is a non-empty, finite collection of these actions. I also omit the braces $\{\}$ outside a singleton $A = \{a\} \in \mathcal{A}$ when this does not lead to confusion.

2.1 *Players and Timeline*

This paper models cognitive dissonance through an intrapersonal game with three stages: $t = 1$, $t = 2$, and an interim stage between them. Three players make decisions sequentially. Players 1 and 2 determine the consumption in each period, while the intermediate player selects a belief for Player 2. Each player has perfect information about the actions taken by their predecessor(s). The timeline of the game is illustrated in Figure 1 below.

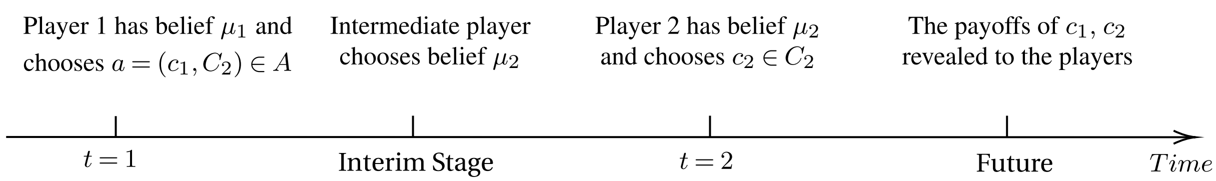


FIGURE 1. Timeline of the Game

The key feature of the CD game is path dependence in decision-making: the period-1 action shapes period-2 choices through belief distortion during the interim stage. Choice influences belief, which in turn shapes future choices. This distortion creates a form of dynamic inconsistency, reflecting the central behavioral deviation from the standard model, as formalized by Axiom 4 in Section 3.2.

2.2 *Payoffs, Strategies and Solution Concept*

Payoffs Let $\mathcal{U}$ denote the set of all non-constant utility indices from X to $\mathbb{R}$. The players share a utility index $u \in \mathcal{U}$. Player 1 and the intermediate player have discount factor $\delta \in [0, 1]$ (discounting is irrelevant for Player 2, as her decision problem is static), and Player 1 has belief $\mu_1 \in \Delta\Omega$. The players' payoff functions are as illustrated in (1) - (3) in the introduction.

Strategies Consider a fixed period-1 menu $A \in \mathcal{A}$. A strategy for Player 1 is a period-1 act $a \in A$. A strategy for the intermediate player is a function $\hat{\mu} : A \to \Delta\Omega$ that specifies a belief for each period-1 action. A strategy for Player 2 is a function $\hat{c}_2 : A \times \Delta\Omega \to \mathcal{C}_2$ such that

[7] For $t = 1, 2$, let $\mathcal{C}_t$ be endowed with the sup metric corresponding to the metric over X. Let $\mathbb{C}_2$ denote the set of period-2 menus endowed with the Hausdorff metric. The set of period-1 actions is endowed with the sup metric over $\mathcal{C}_1 \times \mathbb{C}_2$. The set of period-1 menus is endowed with the corresponding Hausdorff metric.

$\hat{c}_2(a,\mu) \in C_2$ for all $a = (c_1, C_2) \in A$ and $\mu \in \Delta\Omega$. The strategy specifies a consumption c_2 for each period-1 action a and belief μ_2.

Equilibrium Concept Besides the period-1 menu $A \in \mathcal{A}$, the CD game has four parameters: the period-1 belief $\mu_1 \in \Delta\Omega$, the utility index $u \in \mathcal{U}$, the discount factor $\delta \in [0,1]$ and the distortion parameter $\theta \in [0,\infty)$. I use $\kappa = (\mu_1, u, \delta, \theta)$ to refer to the quadruple of these four parameters, with the set of all parameters denoted as K. Next, I define the subgame perfect Nash equilibrium (SPNE) for the CD game with menu $A \in \mathcal{A}$ and parameter $\kappa \in K$.

DEFINITION 2.1. *For $A \in \mathcal{A}$, $\kappa = (\mu_1, u, \delta, \theta) \in K$, the strategy profile $(a^*, \hat{\mu}, \hat{c}_2)$ is a SPNE for the CD game with menu A and parameter κ, if*

$$a^* \in \operatorname*{arg\,max}_{a=(c_1,C_2)\in A} U_{\mu_1}[c_1, \hat{c}_2(a,\hat{\mu})], \tag{4}$$

$$\hat{\mu}(a) \in \operatorname*{arg\,max}_{\mu\in\Delta\Omega} \theta U_\mu[c_1, \hat{c}_2(a,\mu)] - D_{KL}(\mu||\mu_1) \textit{ for all } a = (c_1, C_2) \in A, \tag{5}$$

and

$$\hat{c}_2(a,\mu) \in \operatorname*{arg\,max}_{c_2\in C_2} \mathbb{E}_\mu[u(c_2)] \textit{ for all } a = (c_1, C_2) \in A \textit{ and } \mu \in \Delta\Omega. \tag{6}$$

In the definition above, let $c^* = (c_1^*, c_2^*) \in \mathcal{C}$ denote the consumption stream where for $t \in \mathcal{T}$, $c_t^* \in \mathcal{C}_t$ is the equilibrium period-t consumption in the SPNE. We say c^* is supported by a SPNE in the CD game, denoted by

$$c^* \in SPNE_\kappa(A).$$

The model allows comparative analyses of how the period-1 menu A and the parameter κ influence the equilibrium consumption streams, providing a tool for examining the consequences of cognitive dissonance.[8]

[8] The existence of an SPNE is immediately guaranteed by the assumption that the period-1 and period-2 menus are finite. This existence result can also be generalized to allow for compact menus; in that case, if u is continuous in X, then an application of Harris (1985) implies the existence of an SPNE.

2.3 *Discussion on Cognitive Dissonance*

In the CD game, the internal conflict between one's past action and the belief that it may lead to *undesirable outcomes* leads to cognitive dissonance (Festinger, 1957, Cooper and Fazio, 1984).[9] Dissonance is triggered during the execution of the initial action, as concerns about potential negative outcomes become more *salient* to the DM.[10] The idea that negative outcomes are central to cognitive dissonance is well-established in economics (Akerlof and Dickens, 1982, Rabin, 1994, Konow, 2000, Oxoby, 2003, Mullainathan and Washington, 2009, Chang et al., 2016, Penn, 2017, Acharya et al., 2018, Fan, 2024). Appendix C.1 Table C.1 lists the specific past actions addressed in these studies, along with the corresponding potential negative consequences that lead to cognitive dissonance.

Cognitive dissonance differs from regret, which arises when one realizes a past choice is a mistake. For instance, dissonance can arise when an attractive option is *unavailable*, creating a conflict between one's current preferences and constraints. Regret, on the other hand, stems from recognizing that one failed to choose an attractive option when it was *available*, often accompanied by a sense of missed opportunity. To further illustrate, consider the induced compliance experiments (Festinger and Carlsmith, 1959, Van Veen et al., 2009, Vaidis et al., 2024), where participants initially comply with a requirement to express something they did not believe. Dissonance mitigation then leads them to adjust their beliefs to align with their expressed statements. In these experiments, participants experience dissonance because the desirable option to "express one's true opinion" is *excluded* from the choice set, rather than being *included* in the choice set but ultimately not chosen.[11] Moreover, the participants adjust their opinions "even when no new information is imparted" (Akerlof and Dickens, 1982), indicating that their belief changes are not driven by regret from recognizing a past choice as an *ex-post* mistake.

Despite the differences, the theory of cognitive dissonance and regret minimization can lead to similar predictions – belief distortion may be greater if the initial choice is made

[9] To highlight the role of *undesirable consequences* in triggering cognitive dissonance, Festinger (1957) used the example of a card player who keeps losing while being aware that other players could be professional gamblers. He argued that "this [...] knowledge would be dissonant with his cognition about his behavior, namely, continuing to play. But it should be clear that to specify the relation as dissonant is to assume [the player] wants to win. If for some strange reason this person wants to lose, this relation would be consonant" (p. 13).

[10] Alternatively, dissonance may be triggered because the realization of future outcomes approaches, heightening the agent's worry about potential negative outcomes.

[11] Another option available to participants is to quit the experiment. However, in induced compliance experiments, most (if not all) participants comply with the experimenter's request instead of quitting midway. Consequently, the majority of participants are unlikely to experience significant regret for not quitting.

freely rather than *assigned*. In the cognitive dissonance theory, this is because post-choice dissonance can arise when past behavior suggests incompetence or immorality, undermining one's self-image (Aronson et al., 1968, Akerlof and Dickens, 1982, Harmon-Jones, 2019). Since voluntarily making a poor decision reflects negatively on one's competence, dissonance can be more intense in free-choice scenarios than in assigned-choice scenarios (Chang et al., 2016, Harmon-Jones, 2019). In Section 4.3, I discuss an application of the CD game that captures these patterns, with the DM's types incorporated as one dimension of the state space. High types are more likely to make the correct choice in the short term and earn higher payoffs in the longer term. Therefore, the desire to perceive oneself as a high type creates an incentive to justify the initial choice as superior to the forgone alternatives, which leads to stronger belief distortion in *free-choice* scenarios compared to *assigned-choice* scenarios.

While individuals might adjust their *tastes* rather than *beliefs* over time to align with past actions, I focus on belief adjustment as a channel for mitigating dissonance. As shown in Section 4, taking this approach allows the model to capture various empirical regularities documented in the cognitive dissonance literature.

Other forms of inconsistency may also trigger dissonance. For example, a DM may feel uncomfortable when new information contradicts existing beliefs, leading to confirmation bias (Yariv, 2005) or information avoidance (Harmon-Jones and Mills, 2019). Chauvin (2021) argues that dissonance can arise when two pieces of information conflict with each other. Wang (2022) studies how regret aversion can lead to information avoidance. These phenomena, however, are beyond the scope of this study.

Section 6 explains the rationale behind other assumptions in the CD game, including Player 1's rational expectations, the forward-looking behavior of the intermediate player, and the choice of the cost function for belief distortion.

3. Characterization

This section characterizes the cognitive dissonance game in a framework *à la* Anscombe and Aumann (1963). Let Z be a finite set of outcomes. In this section, I assume the payoff space X is ΔZ, the set of lotteries over the outcomes. Let $\mathcal{U}^*$ define the collection of all non-constant linear utility functions over X. I characterize the equilibria of CD games with parameter κ, where $\kappa = (\mu, u, \delta, \theta) \in K^* \equiv Int(\Delta\Omega) \times \mathcal{U}^* \times (0,1) \times [0,\infty)$. Section 3.1 discusses the choice data. Section 3.2 introduces the axioms. Section 3.3 shows the main representation theorem and the uniqueness result. Section 3.4 introduces the closed-form solution to the CD game. Section 3.5 proposes a comparative notion for belief distortion induced by cognitive dissonance.

3.1 *Primitives*

For period-1 action $a = (c_1, C_2)$, the set of *feasible consumption streams* is denoted by $F(a) = \{(c_1, c_2) \mid c_2 \in C_2\}$. For period-1 menu A, the set of feasible consumption streams is $F(A) = \cup_{a \in A} F(a)$. The central mathematical object in my analysis is an induced choice correspondence, which assigns each period-1 menu A to a set of feasible consumption streams. The formal definition is provided in Definition 3.1 below.

DEFINITION 3.1. *An induced choice correspondence is a non-empty correspondence* $\gamma : \mathcal{A} \rightsquigarrow \mathcal{C}$ *with* $\gamma(A) \subset F(A)$ *for every* $A \in \mathcal{A}$*. Additionally, for* $t \in \mathcal{T}$ *and* $\omega \in \Omega$*, define* $\gamma_{t\omega}(A) = \{c_{t\omega} \mid c \in \gamma(A)\}$.

In principle, the choice data available to the modeler maps each period-1 menu to a set of triplets (c_1, C_2, c_2), where (c_1, C_2) represents the choice in period 1, and c_2 represents the choice in period 2. Definition 3.1 instead uses a "thinner" dataset that maps each period-1 menu to the consumption stream (c_1, c_2) resulting from choices in both periods. This approach simplifies the notation, making the axioms clearer and easier to interpret.

Definition 3.2 introduces two key classes of period-1 menus that are important in characterizing the CD game.

DEFINITION 3.2. *Let* $A \in \mathcal{A}$.

- A *is 1-determined, denoted by* $A \in \mathcal{A}_1$*, if* $|C_2| = 1$ *for all* $(c_1, C_2) \in A$.
- A *is 2-determined, denoted by* $A \in \mathcal{A}_2$*, if* $|A| = 1$.

If A is *1-determined*, the consumption stream $\gamma(A)$ is entirely determined by the choice at time $t = 1$, as the period-2 menu C_2 contains only one option. On the other hand, if A is *2-determined*, the consumption stream $\gamma(A)$ is entirely determined by the choice at time $t = 2$, as the period-1 menu A contains only one element. As we will see in Section 3.2, analyzing the behavior of γ over $\mathcal{A}_1$ and $\mathcal{A}_2$ provides the modeler useful information to understand the decision-making process in both periods.

Instead of choice correspondences, one might consider modeling the period-1 and period-2 choices using preference relations. However, this interpretation can confound the understanding of CD games for two reasons. First, Player 1's choices reflect not only her own "preferences" but also the strategies of subsequent players. Second, it is difficult to interpret Player 2's choices as generated by her "preference" in the traditional sense because her belief, determined by the intermediate player, is menu-dependent.

3.2 *Axioms*

This section introduces the axioms characterizing the cognitive dissonance game. I first characterize the rational benchmark of time-separable SEU in the context of my framework (Theorem 3.1) using four axioms: Axioms 1- 3 and dynamic consistency (Axiom DC). To characterize the cognitive dissonance game, I demonstrate how dynamic consistency must be relaxed in a specific manner. Theorem 3.2 then presents a complete characterization of the CD game.

I start with introducing Axiom 1, which contains variants of the Anscombe and Aumann (1963) postulates that guarantee standard properties for decision-making. The axiom applies to a restricted domain, $\mathcal{A}_1 \cup \mathcal{A}_2 \subset \mathcal{A}$, which includes all 1-determined and 2-determined menus as defined in Definition 3.2 above. I introduce the following notation to facilitate the statement of Axiom 1. First, for $t \in \mathcal{T}$ and $A, A', B \in \mathcal{A}_t$, we say $A' = A \sqcap B$ if $F(A') = F(A) \cap F(B)$. Second, for finite $L \subset X$, $c, c' \in \mathcal{C}$, $t \in \mathcal{T}$ and $\omega \in \Omega$, we say $c' \in L^c_{t\omega}$ if for all $s \in \mathcal{T}$ and $\psi \in \Omega$,

$$c'_{s\psi} \in \begin{cases} L & \text{if } s = t, \psi = \omega; \\ \{c_{s\psi}\} & \text{otherwise.} \end{cases}$$

The DM's choice from $L^c_{t\omega}$ affects her payoff only at time t and when the true state is ω; for other times and states, the payoff follows the predetermined consumption stream c.

Axiom 1. *(Standard Choice) For $\alpha \in (0,1)$, $\tau \in \mathcal{T}$, $A, B \in \mathcal{A}_\tau$, and a finite $L \subset X$,*

(i) (WARP) if $c, c' \in \gamma(A \sqcap B)$ and $c \in \gamma(A)$, then $c' \in \gamma(A)$;

(ii) (Non-Degeneracy) there exists $A^* \in \mathcal{A}_1$ such that $\gamma(A^*) \neq F(A^*)$;

(iii) (Independence) if $A, B \in \mathcal{A}_1$, then $\gamma(\alpha A + (1-\alpha)B) = \alpha\gamma(A) + (1-\alpha)\gamma(B)$;

(iv) (Continuity) γ is upper hemicontinuous in $\mathcal{A}_1 \cup \mathcal{A}_2$;

(v) (State Independence) for $s, t \in \mathcal{T}$, $\omega, \psi \in \Omega$, and $c, c' \in \mathcal{C}$, $\gamma_{t\omega}(L^c_{t\omega}) = \gamma_{s\psi}(L^{c'}_{s\psi})$.

In the restricted domain $\mathcal{A}_1 \cup \mathcal{A}_2$, Axiom 1 outlines five properties akin to the Anscombe and Aumann (1963) axioms for SEU. The framing of Axiom 1 is inspired by Axiom 2.1 (Standard Receiver Preferences) from Jakobsen (2021), which formulates the Anscombe-Aumann axioms within the context of choice correspondences. All five assumptions apply to $\mathcal{A}_1$, which implies the period-1 choices can be rationalized by a time-separable SEU function taking the form

$$\mathbb{E}_\mu[u(c_1)] + \delta \mathbb{E}_{\mu'}[u(c_2)], \tag{7}$$

where $\mu, \mu' \in \Delta\Omega$. Two points are important for understanding the constraints imposed by Axiom 1 for period-2 choices. First, (iii) (akin to the classic *Independence* axiom) applies only to $\mathcal{A}_1$, not $\mathcal{A}_2$, and hence is not assumed for period-2 choices. I need to weaken Independence this way to accommodate the model because the period-2 belief is adjusted to mitigate the cognitive dissonance from the previously chosen menu C_2, generating violations of Independence. Second, (i), (iv), and (v), corresponding to the weak axiom of revealed preference (WARP), continuity, and state independence, are retained in $\mathcal{A}_2$ to minimize deviations from the standard model.

AXIOM 2. *(Impatience) If* $\gamma(\{(c_1, c_1), (d_1, d_1)\}) = (c_1, c_1)$*, then* $\gamma(\{(c_1, d_1), (d_1, c_1)\}) = (c_1, d_1)$.

Axiom 2 states that the period-1 DM is impatient. If a consumption stream (c_1, c_1) is chosen over (d_1, d_1), then the DM would prefer to consume c_1 earlier than d_1. In the SEU representation in (7), Axiom 2 also guarantees that $\mu = \mu'$, reflecting that there is only *one* draw of the true state before the start of the game.

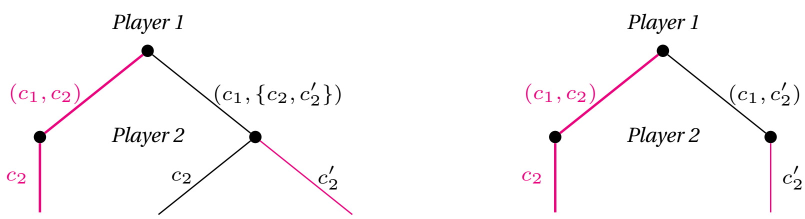
DEFINITION 3.3. *For induced choice correspondence* γ*, menu* $A' = \{(c_1^i, c_2^i)\}_{i \in I}$ *is a consistent selection from menu* $A = \{(c_1^i, C_2^i)\}_{i \in I}$ *if for all* $i \in I$*,*

$$(c_1^i, c_2^i) \in \gamma[(c_1^i, C_2^i)]. \tag{8}$$

In the context of the CD game, a consistent selection refines the period-1 decision problem A by removing all off-path decision nodes in each subgame. For each $i \in I$, once the off-path nodes are eliminated, the remaining consumption stream c^i reflects Player 2's equilibrium choice in the subgame corresponding to the period-1 action a^i.

AXIOM 3. *(Sophistication)* $c \in \gamma(A)$ *if and only if* $c \in \gamma(A')$ *for a consistent selection* A' *from* A.

Axiom 3 states that the period-1 DM is sophisticated and uses backward induction to guide her choices. Consider period-1 menu $A = \{c, a\}$ where $c = (c_1, c_2)$ and $a = (c_1, \{c_2, c_2'\})$, as depicted in Figure 2. Although the period-1 self would choose c_2 over c_2', she anticipates that her future self will exhibit dynamic inconsistency and select c_2' instead. Therefore, $\gamma(\{(c_1,c_2),(c_1,\{c_2,c_2'\})\}) = \gamma(\{(c_1,c_2),(c_1,c_2')\}) = (c_1,c_2)$. This behavior aligns with the backward induction in consistent planning models pioneered by Strotz (1955), which is also termed "folding back" (Machina, 1989).



(a) $\gamma(A) = c$ (b) $\gamma(A') = c$

FIGURE 2. Decision Trees in Axiom 3

AXIOM DC. *(Dynamic Consistency) For period-1 action* a, $\gamma[F(a)] = c^* \implies \gamma(a) = c^*$.

Axiom DC states that if the period-1 self would choose c^* with the ability to commit to any feasible consumption stream (denoted by $F(a)$), then the period-2 self should also choose c^*. As illustrated in Theorem 3.1 below, Axiom 1, 2, 3 and DC jointly characterize the standard model of time-separable SEU, which corresponds to the CD game with dissonance factor $\theta = 0$.

THEOREM 3.1. $\gamma : \mathcal{A} \rightsquigarrow \mathcal{C}$ *satisfies Axioms 1, 2, 3, and DC if and only if there exists* $\kappa = (\mu, u, \delta, \theta) \in K^*$, *where* $\theta = 0$, *such that for every* $A \in \mathcal{A}$,

$$\gamma(A) = SPNE_\kappa(A).$$

Moreover, μ, δ *are unique, and* u *is unique up to affine transformations.*

The goal of the rest of my analysis is to find a suitable weakening of Axiom DC that characterizes the CD game for the general case where the dissonance factor $\theta \geq 0$. A key element of this characterization is formalizing when an event is considered *desirable* for a period-1 action a. This is central to understanding cognitive dissonance, which arises from the discomfort caused by the possibility that a past action might lead to *undesirable* outcomes.

Definition 3.4 formally states when an event E is considered desirable. To facilitate the definition, for an event $E \in \mathcal{E}$, period-1 action a is E-*measurable* if, for every feasible consumption stream $c \in F(a)$ and $t \in \mathcal{T}$, $c_t : \Omega \to X$ is measurable with respect to the partition $\{E, E^c\}$. In this context, define the payoff stream in E corresponding to c by $c_E = (c_{1\omega^*}, c_{2\omega^*})$ for some $\omega^* \in E$.

DEFINITION 3.4. *An event $E \in \mathcal{E}$ is desirable for period-1 action a if a is E-measurable and for any $c \in F(a)$,*

$$c_E \in \gamma(\{c_E, c_{E^c}\}).$$

Recall that c_E (*resp.* c_{E^c}) defines the payoff stream in event E (*resp.* E^c). Definition 3.4 states that, for any feasible consumption stream c, the corresponding payoff stream in E is always "preferred" over that in E^c. More precisely, if the DM could control the realization of E and E^c, she would choose E.

Next, I introduce Axiom 4, the key axiom for characterizing the CD game. For any $E \in \mathcal{E}$ and any E-measurable action a, let $F_E(a) = \{c_E \mid c \in F(a)\}$ denote the set of feasible payoff streams in event E.

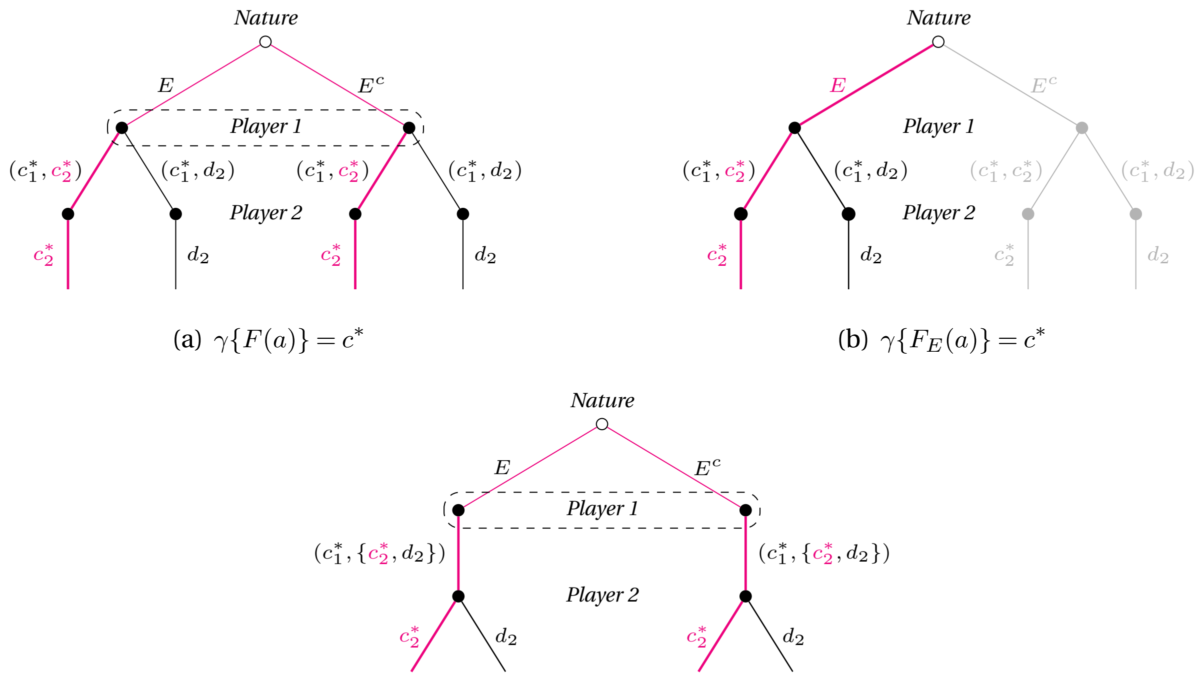
AXIOM 4. *(Cognitive Consistency) For $E \in \mathcal{E}$ and period-1 action a, if E is desirable for a, then*

$$\begin{cases} \gamma[F(a)] & = c^* \\ \gamma[F_E(a)] & = c^*_E \end{cases} \implies \gamma(a) = c^*. \tag{9}$$

Axiom 4 weakens Axiom DC. It states that the DM would choose c^* if this option remains optimal *regardless* whether the DM succumbs to the desire to alleviate cognitive dissonance.

- "$\gamma[F(a)] = c^*$" indicates that, when unaffected by cognitive dissonance, the *ex-ante* self considers c^* as the optimal choice. This is illustrated in part (a) of Figure 3.
- "$\gamma[F_E(a)] = c^*_E$" states that, even if the DM fully surrenders to the desire to reduce dissonance by convincing herself that the favorable event E will surely occur, she still sees c^* as optimal. This is shown in part (b) of Figure 3.

In this case, the period-2 self (Player 2 in the CD game) has no incentive to deviate from c^*. Formally, $\gamma(a) = c^*$. This conclusion is represented in part (c) of Figure 3. By weakening Axiom DC, Axiom 4 captures two sources of cognitive dissonance: past actions that influence future payoffs, as well as past commitments that bind future choices.



(a) $\gamma\{F(a)\} = c^*$

(b) $\gamma\{F_E(a)\} = c^*$

(c) $\gamma(a) = c^*$

FIGURE 3. Decision Trees in Axiom 4

Past Action Influencing Future Payoffs Consider Example 1, where the period-1 choice is $a = (c_1, \{c_2, \emptyset_2\})$ with c_t (*resp.* $\emptyset_t$) representing working (*resp.* not working) at the chemical plant in $t = 1, 2$. Event $\{G\}$ represents a safe workplace, which is favorable for a as it is always weakly better than a poisonous workplace. For the period-1 self, $\gamma[F(a)] = (c_1, \emptyset_2)$, however, $\gamma\left[F_{\{G\}}(A)\right] \neq (c_{1G}, \emptyset_{2G})$ as the payoff of $\emptyset_2$ in G equals 0, which is worse than the 2 offered by c_2. In this case, Axiom 4 allows the period-2 self to deviate to c_2, consistent with the behavior in Example 1.

Past Commitment Binding Future Choices Consider a modified version of Example 1 inspired by Akerlof and Dickens (1982). The period-1 action is $a' = (\emptyset_1, \{c_2, c_2'\})$, where for each period, $\emptyset_t, c_t$ are defined as above, and c_2' represents the use of a safety protection measure that partially mitigates health risks, offering a payoff of 1 if the chemical is safe (G) and -1 if it is poisonous (B). This period-1 action can be interpreted as signing a contract that commits to $\{c_2, c_2'\}$, which poses no immediate health risk but leads the worker to anticipate future chemical exposure, raising safety concerns and cognitive dissonance. For the period-1 self, it is optimal to use safety protection in period 2, that is, $\gamma[F(a')] = (\emptyset_1, c_2')$.[12] However, $\gamma\left[F_{\{G\}}(a')\right] \neq (\emptyset_{1G}, c'_{2G})$ as the payoff of c_2' in G equals 1, which is worse than the 2 offered by c_2. In this case, Axiom 4 allows the DM to deviate to using no safety protection (c_2), consistent with the neglect of risks in the workplace induced by dissonance mitigation.

Next, I present Axiom 5 and 6, two standard axioms that give more structure to the CD game. For $E \in \mathcal{E}$, period-1 action $a = (c_1, C_2)$ and consumption stream $d = (d_1, d_2)$, I define $aEd = (c_1Ed_1, \{c_2Ed_2 \mid c_2 \in C_2\})$, where for $t \in \mathcal{T}$ and $\omega \in \Omega$, $[c_tEd_t] \in \mathcal{C}_t$ is the consumption with state-ω payoff equal to $c_{t\omega}$ if $\omega \in E$ and $d_{t\omega}$ if $\omega \notin E$.

AXIOM 5. *(STP) For every $E \in \mathcal{E}$ and period-1 action a, $cEc' \in \gamma(aEc') \implies cEc'' \in \gamma(aEc'')$.*

Axiom 5 is a variant of Savage's P2 (Sure-thing principle, or STP). It states that if the payoffs of the consumption streams differ only in the states in E (agreeing otherwise), the chosen consumption stream should be determined solely by their differences in E. I include Axiom 5 in my axiomatic system for two main reasons. First, it rules out various violations of subjective expected utility theory caused by factors other than cognitive dissonance, such as the behavioral pattern observed in the one-urn Ellsberg paradox. Second, an important extension of my model considers belief updating in the interim stage. In this context, STP would guarantee that conditional choices are well-defined using the standard definition of conditional preference (Ghirardato, 2002). Specifically, for each event $E \in \mathcal{E}$ realized in the interim stage, Axiom 5 ensures that the conditional choice correspondence $\gamma^E : \mathcal{A} \rightsquigarrow \mathcal{C}$ is well-defined, where $c \in \gamma^E(A)$ if and only if, for all $c' \in \mathcal{C}$,

$$cEc' \in \gamma(\{aEc' \mid a \in A\}).$$[13]

AXIOM 6. *(Weak C-Independence) For $a = (c_1, C_2)$, $x, y \in X$ and $\alpha \in (0,1)$,*

$$(c_1, \alpha c_2 + (1-\alpha)x) \in \gamma[(c_1, \alpha C_2 + (1-\alpha)x)] \implies (c_1, \alpha c_2 + (1-\alpha)y) \in \gamma[(c_1, \alpha C_2 + (1-\alpha)y)].$$

[12]Notice that $c_2 = (2, -4)$ and $c_2' = (1, -1)$, while the period-1 self has an equal prior over both states. As a result, the expected payoff of c_2' is 0, which is higher than the expected payoff of -1 given by c_2.

[13]For $A \in \mathcal{A}$ and $c \in \mathcal{C}$, define $AEc = \{aEc \mid a \in A\} \in \mathcal{A}$. To prove that γ^E is well-defined, I fix $A \in \mathcal{A}$ and show that $\gamma(A) \neq \emptyset$. It suffices to show that for $c', c'' \in \mathcal{C}$, $cEc' \in \gamma(AEc') \iff cEc'' \in \gamma(AEc'')$. To prove this, notice that by Axiom 5, for all $a \in A$, $cEc' \in \gamma(aEc')$ if and only if $cEc'' \in \gamma(aEc'')$. Take $A' \in \mathcal{A}$ such that $A'Ec'$ is a consistent selection of AEc', then $A'Ec''$ must be a consistent selection of AEc''. Since the period-1 self is an SEU maximizer, $cEc' \in \gamma(A'Ec') \iff cEc'' \in \gamma(A'Ec'')$. Axiom 3 then implies $cEc' \in \gamma(AEc') \iff cEc'' \in \gamma(AEc'')$, the desired result.

Recall from the discussion on Axiom 1 that cognitive dissonance mitigation implies that beliefs in period 2 are menu-dependent, leading to violations of *Independence* for choices in period 2. To account for this, Axiom 6 asserts that the choices in the second period adhere to Weak C-Independence (Maccheroni et al., 2006). Assuming the other axioms hold, strengthening the axiom to C-Independence (Gilboa and Schmeidler, 1989) or Independence will restore the standard model of time-separable SEU.

3.3 *Representation Theorem*

The following theorem characterizes cognitive dissonance games with Axioms 1-6.

THEOREM 3.2. *$\gamma : \mathcal{A} \rightsquigarrow \mathcal{C}$ satisfies Axioms 1-6 if and only if there exists $\kappa \in K^*$ such that for every $A \in \mathcal{A}$,*

$$\gamma(A) = SPNE_\kappa(A).$$

I denote this induced choice correspondence by γ_κ. For parameters $\kappa = (\mu_1, u, \delta, \theta)$ and $\kappa' = (\mu_1', u', \delta', \theta') \in K^*$, $\gamma_\kappa = \gamma_{\kappa'}$ if and only if $\mu_1 = \mu_1'$, $\delta = \delta'$, and $\theta u = \theta' u' + b$ for some $b \in \mathbb{R}$.

Theorem 3.2 states that if Axioms 1-6 hold, the DM behaves as if multiple selves sequentially move in the CD game, characterized by a parameter κ. Conversely, the equilibrium consumption streams of CD games with parameter κ must satisfy all the axioms. Furthermore, the parameters exhibit standard uniqueness properties and are identifiable from choice data.

3.4 *Solving the CD game*

Fix a period-1 action $a = (c_1, C_2)$ and consider the corresponding subgame starting at the interim stage. I introduce Proposition 3.1, which characterizes the set of equilibrium period-2 consumptions in the subgame. An important role will be played by the class of transformations ϕ_θ, indexed by $\theta \in [0, \infty)$:

$$\phi_\theta(u) = \begin{cases} \exp(\theta u) & \text{for } \theta > 0, \\ u & \text{for } \theta = 0. \end{cases} \tag{10}$$

Proposition 3.1. *For $\kappa = (\mu_1, u, \delta, \theta) \in K^*$ and period-1 action $a = (c_1, C_2)$, the consumption stream $(c_1, c_2^*) \in \gamma_\kappa(a)$ if and only if*

$$c_2^* \in \arg\max_{c_2 \in C_2} \mathbb{E}_{\mu_1}[\phi_\theta(u(c_1) + \delta u(c_2))]. \tag{11}$$

Proposition 3.1 is central to solving the CD game. For $A \in \mathcal{A}$, we may use (11) to solve for the period-2 equilibrium consumption and solve the period-1 choices by backward induction. Substituting the equilibrium actions into the intermediate player's optimization problem gives the equilibrium period-2 belief. An important feature of the utility function in (11) is *intertemporal complementarity*, which is also the key feature for habit formation models (Constantinides, 1990). To see why, notice that when $\theta > 0$,

$$\frac{\partial^2 \phi_\theta(u + \delta v)}{\partial u \partial v} = \delta \phi_\theta(u + \delta v) > 0.$$

Unlike many habit formation models, the CD game emphasizes how past actions influence future decisions by distorting beliefs. Sections 4 and 5 discuss how cognitive dissonance helps explain patterns of path dependency in decision-making that habit formation models do not fully capture.

3.5 *Comparative Cognitive Dissonance*

This section develops a comparative notion of the degree to which a DM is susceptible to cognitive dissonance. If someone is more influenced by cognitive dissonance, her belief would be distorted more significantly after the choice in the first period, and the modeler should observe more frequent violations of Axiom DC. To formalize this idea, for period-1 action a, I say the induced choice correspondence γ is a-*DC* if $\gamma[F(a)] = \gamma(a)$.[14]

Definition 3.5. *For induced choice correspondences γ, γ', we say γ is more CD-susceptible than γ' if for event $E \in \mathcal{E}$,*

- *$\gamma(A) = \gamma'(A)$ for all $A \in \mathcal{A}_1$, and*
- *if E is desirable for period-1 action a, then*

$$\gamma \text{ is } a\text{-DC} \implies \gamma' \text{ is } a\text{-DC}.$$

[14] Similar to Axiom DC, $\gamma[F(a)]$ represents the period-1 self's choices when she can commit to any feasible consumption stream, while $\gamma(a)$ represents the choices made by the period-2 self.

First, Definition 3.5 states that the choices are identical for γ and γ' if $A \in \mathcal{A}_1$, where the period-1 self has full commitment power and is the sole controller of the consumption stream. Thus, differences between γ and γ' should only arise from variations in the degree of cognitive dissonance. Second, Definition 3.5 considers the case where event E is desirable, where cognitive dissonance causes violations of DC by leading the DM to overestimate the likelihood of favorable outcomes. In this case, γ is "less" dynamically consistent than γ': whenever the former exhibits DC, the latter must also exhibit DC. The choices in γ thus reflect more significant belief distortions due to dissonance mitigation.

THEOREM 3.3. *For $\kappa = (\mu_1, u, \delta, \theta)$ and $\kappa' = (\mu_1', u, \delta', \theta') \in K^*$, the following are equivalent:*

- *γ_κ is more CD-susceptible than $\gamma_{\kappa'}$;*
- *$\mu_1 = \mu_1'$, $\delta = \delta'$, and $\theta \geq \theta'$.*

Theorem 3.3 provides a behavioral interpretation of the parameter θ: a larger θ implies more frequent occurrences of dynamic inconsistency, driven by the mitigation of cognitive dissonance.

4. ACCOMMODATING EMPIRICAL REGULARITIES

In this section, I first outline several key stylized facts from the cognitive dissonance literature, along with the empirical works supporting each of them. I then present a single application of the CD game consistent with these stylized facts. Finally, in Section 4.3, I propose an extension of the CD game that accounts for more insights from the cognitive dissonance literature.

4.1 *Stylized Facts*

I start with effort justification and induced compliance, two of the foundational elements in the cognitive dissonance literature (Harmon-Jones and Mills, 2019).

Effort Justification People experience cognitive dissonance after they invest heavily in a task that turns out to be unworthy. Dissonance reduction thereby leads people to engage in *effort justification* to convince themselves that the task was more valuable than they had previously thought (Harmon-Jones and Mills, 2019).[15] Multiple pieces of evidence in the literature align with effort justification. For example, people tend to view a group more positively after undergoing a challenging initiation process (Aronson and Mills, 1959) and develop a stronger belief in a horse's likelihood of winning after spending money to place a bet on it (Knox and Inkster, 1968). Similarly, individuals who invest more effort into pro-social behaviors are more likely to continue such actions (Gneezy et al., 2012), and investors often hold onto depreciating assets to justify previous investment decisions (Staw, 1981, Chang et al., 2016). Effort justification provides a psychological foundation for the sunk-cost fallacy – the tendency to persist in an endeavor due to prior investments, even when quitting would lead to better outcomes (Eyster et al., 2022).

Induced Compliance In effort justification, individuals initially invest time, energy, or resources in pursuing an option they perceive as desirable. However, if the costs associated with this desirable option become sufficiently high, they may feel pressured to select an alternative that contradicts their judgment. *Induced compliance* refers to the phenomenon where, after being pressured to perform an action, individuals change their opinions in favor of it (Harmon-Jones and Mills, 2019). This aligns with laboratory evidence showing that participants, after being instructed to express something they do not initially believe, tend to adjust their beliefs to match their statements (Festinger, 1957, Van Veen et al., 2009, Vaidis et al., 2024).[16] This explains why people under oppression may internalize imposed beliefs as truth (Jost and Banaji, 1994, David, 2013). For example, a citizen living under media censorship may find it more comfortable to believe that the censorship is beneficial (Yang, 2022), the propaganda provides useful information (Shen, 2022), or that the blocked sources spread disinformation.[17] Similarly, women may self-stereotype, believing the quality of their contribution is lower than males (Jost, 1997), especially in areas that are stereotypically outside of their gender's domain (Coffman, 2014).

[15] Effort justification can also lead individuals to embrace the "just-world hypothesis" (Benabou and Tirole, 2006), fostering the belief that personal effort will pay off in the long run, while attributing misfortune to a lack of effort or laziness.

[16] The experimenter's instructions create social pressure, deterring participants from expressing their true opinions.

[17] Yang (2022) provides laboratory evidence linking justification in a censorship system to cognitive dissonance mitigation.

Negative-Incentive Effect Offering compensation to encourage compliance can boost adherence, yet it may also reduce the degree of opinion shift in favor of the action—a phenomenon known as the negative-incentive effect. Cognitive dissonance theory suggests that substantial compensation eases dissonance by reducing concerns about negative outcomes and strengthening the rationale for adherence, which, in turn, lessens motivation for opinion shifts. Consistent with the negative-incentive effect, experiments show that participants who are paid *more* to speak positively about a task are less likely to believe the task is "interesting" or "scientifically important" *ex-post* (Festinger and Carlsmith, 1959, Van Veen et al., 2009). Similarly, a field experiment by Meier (2007) found that subsidizing charitable donations can ultimately cause donation levels to drop below those observed in the pre-subsidy period. Cognitive dissonance explains this observation by suggesting that belief distortion in favor of donation behavior weakens when the donation is subsidized.The negative-incentive effect suggests that lasting changes in behaviors may be better achieved with a *nudge* that is "just enough... to elicit overt compliance" (Festinger, 1957, p. 95), rather than with large rewards or punishments.

Dissonance-Driven Polarization Dissonance mitigation leads people to have more confidence in the option they chose compared to the unchosen one (Harmon-Jones and Mills, 2019). For example, in a classical lab experiment, after voluntarily choosing a product, people tend to view the chosen item as more useful than the one they passed up (Brehm, 1956). This effect can heighten disagreements among individuals who made different past choices based on differing opinions (Acharya et al., 2018). For instance, Mullainathan and Washington (2009) empirically shows that simply voting for a specific candidate in an election can strengthen future confidence in that candidate, contributing to greater polarization among voters compared to non-voters.

4.2 *Predictions of the CD Model*

In this section, I study how temporary external pressure can lead to a persistent shift in beliefs, thereby altering long-term behavior. My results mirror the stylized facts discussed above. To illustrate, I draw on an example inspired by the concepts of cognitive dissonance influencing political choice (Mullainathan and Washington, 2009, Acharya et al., 2018). Later, I will show how my setup can also connect directly with the specific empirical evidence referenced earlier.

A DM chooses the degree to which she supports a political party. She perceives uncertainty over which party, L or R, is a better choice, represented by the binary state $\Omega = \{L, R\}$. Besides the state of the world, the DM's payoff depends on two factors: $s \in S$, representing her choice of political stance, and $\chi \geq 0$, an exogenous parameter that represents the *social pressure* within her social circle that favors supporters of L and punishes R. Formally, for each political choice s and pressure level χ, the consumption $c(s, \chi)$ is given by $(-s - \chi s, s - \chi s)$, where the two entries correspond to the payoffs in state L and R respectively. Choosing a $s > 0$ reflects support for party R, which incurs a penalty from social pressure. In contrast, choosing $s < 0$ reflects supporting party L, which aligns with the social pressure and is thus rewarded. For pressure level χ, I define the induced feasible consumptions as $C(\chi) = \{c(s, \chi) \mid s \in S\}$. The agent faces the period-1 menu $A(\chi) = C(\chi) \times \{C(0)\}$, making separate political choices for each period, with the pressure χ present only at time $t = 1$ and removed by time $t = 2$. I employ this setting to analyze how temporary external pressure can cause lasting shifts in beliefs. Assume that $u'' < 0$ and $\theta > 0$.

I analyze the comparative statics of the equilibrium outcomes with respect to the initial external pressure χ, holding the parameters for the CD game fixed. For each $\chi \in \mathcal{X}$, $\mu_2^*(\chi)$ is the collection of all equilibrium period-2 beliefs in state R, while $s_t^*(\chi)$ is the collection of all equilibrium period-t political choice ($t = 1, 2$).

ASSUMPTION 1. *$S = \{l, r\}$, where $l < 0 < r$. Moreover, $s_1^*(0) = s_2^*(0) = r$.*

Assumption 1 states that the set of feasible political choices includes an option $l < 0$ (supporting party L) and another option $r > 0$ (supporting party R). Furthermore, the DM's confidence in R is strong enough for her to consistently choose r without any external influence. Based on this assumption, I present a proposition demonstrating how the initial pressure χ can have a lasting impact on the agent's belief, capturing the effects of effort justification, induced compliance, and negative incentives discussed previously.

PROPOSITION 4.1. *There exists a threshold pressure χ^* such that $s_1^*(\chi) = r$ if $\chi < \chi^*$ and $s_1^*(\chi) = l$ if $\chi > \chi^*$. Moreover,*

- *for the interval $I = [0, \chi^*)$ or (χ^*, ∞), $\mu_2^*(\chi)$ is strictly increasing within I;*
- *for the threshold pressure χ^*, it holds that*

$$\lim_{\chi \uparrow \chi^*} \mu_2^*(\chi) > \lim_{\chi \downarrow \chi^*} \mu_2^*(\chi).$$

Proposition 4.1 establishes a non-monotonic pattern in $\mu_2^*(\chi)$, the equilibrium period-2 belief that R is the better political party, given the initial social pressure χ. Figure 4 below illustrates this pattern.[18] The horizontal axis represents χ, while the vertical axis shows the logarithm of the ratio between the period-2 beliefs in R and L. As period-1 pressure χ against supporting R increases, period-2 confidence in R initially rises with χ, drops sharply at a threshold χ^*, and then resumes increasing beyond this point.

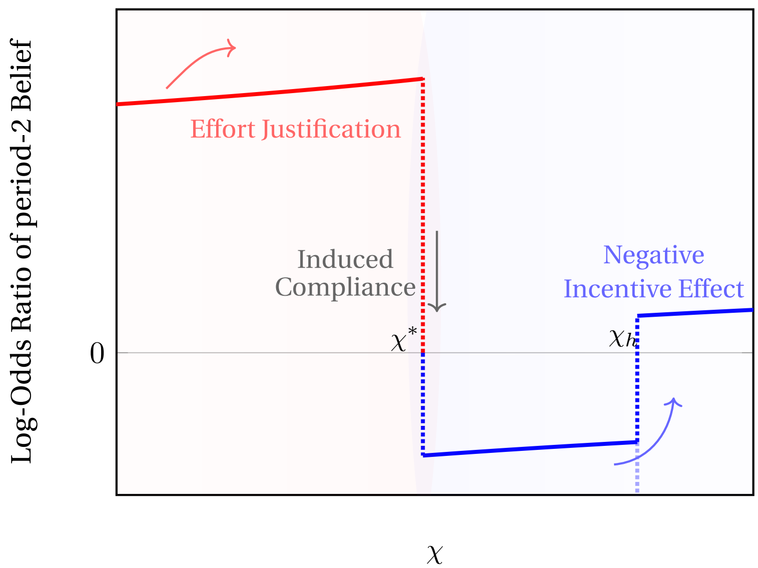


FIGURE 4. Initial Pressure and Evolution of Beliefs

Effort Justification Consider $\chi \in [0, \chi^*)$. In this range, the DM chooses r in the first period, supporting party R. As χ increases, she must expend more effort—whether in time, energy, or resources—to resist the social pressure and maintain her support for R. This sacrifice intensifies her need to justify her choice, strengthening her belief in R by period 2. In this case, the increased pressure to discourage support for party R strengthens the agent's commitment to it.

Induced Compliance At the threshold χ^*, the DM faces social pressure strong enough to compel her to switch from supporting R to publicly supporting L. To reduce discomfort from this switch, her period-2 belief in R drops discontinuously.

[18] Figure 4 is generated with a calibration of $\theta = 0.5$, $\delta = 0.5$, $u = \ln(x+1)$, $\mu_{1R} = 0.6$, $l = -0.6$, and $r = 0.5$.

Negative-Incentive Effect Consider $\chi \in (\chi^*, \infty)$. In this region, social pressure prompts the DM to support L in the first period. As χ continues to increase, the social rewards for supporting L grow, mitigating cognitive dissonance from this shift. This reduction in dissonance dampens belief distortion favoring L, and as a result, confidence in R resumes increasing.

PROPOSITION 4.2. *If $\chi = 0$, then the equilibrium period-2 belief in R is strictly greater than the period-1 belief in R.*

Proposition 4.2 accounts for the *polarization* caused by cognitive dissonance mitigation. Without external pressure, the DM's confidence in R increases over time, driven by the need to justify the choice of r. To clarify the connection to polarization, consider the opposite scenario: if initial confidence in R is sufficiently low, the DM chooses l in both periods without external pressure. In this case, the desire to justify l strengthens confidence in state L over time. This moves opposite to Proposition 4.2, reflecting polarization.

REMARK 1. The setup can be modified appropriately to align closely with the empirical evidence I discussed in Section 4.1. For example, consider $s \in S = \{0, 1\}$, where $s = 1$ represents making an investment in a broad sense, which may correspond to expending effort in an initiation process (Aronson and Mills, 1959), betting on a horse (Knox and Inkster, 1968), donating money to others (Gneezy et al., 2012), or investing in a financial asset, as in Chang et al. (2016). The parameter χ represents the effort (money, time, energy...) the DM needs to pay to make the investment. The state R represents that the investment turns out to be valuable for the DM, while L represents the contrary.

REMARK 2. While my analysis in Proposition 4.1 and 4.2 focus on period-2 *beliefs*, these result have *choice* implications. For example, due to the negative-incentive effect, the equilibrium period-2 choice can shift from l to r as external pressure increases in the range (χ^*, ∞). Specifically, in Figure 4, this switch occurs when pressure rises from just below χ_h to just above it. This switch is internalized by the intermediate player in the CD game, leading to a discontinuity at χ_h in the figure.[19]

[19] My setup considers $S = \{l, r\}$ where $l < r$, which does not allow the DM to switch to options greater than r. I may extend this setting by considering $S = \{l, r, \bar{r}\}$, where $l < 0 < r < \bar{r}$. In this case, all results in this section still hold; however, with the addition of $\bar{r}$, effort justification could lead the DM to escalate support from r to the stronger option $\bar{r}$ in the second period.

4.3 *Cognitive Dissonance and Self Image*

This section introduces an extension of the CD game to capture an additional aspect of cognitive dissonance.

When a choice is made *freely* rather than *assigned* by someone else, a poor choice can appear as a sign of incompetence, thereby intensifying post-choice dissonance (Harmon-Jones and Mills, 2019). For example, investors are more reluctant to admit a past investment was a mistake when they made the decision themselves, compared to when it was delegated to a mutual fund manager (Chang et al., 2016). In laboratory settings, despite some recent challenges (Vaidis et al., 2024),[20] psychologists typically find that reminding participants of their *voluntary* participation intensifies cognitive dissonance compared to situations where participation is perceived as mandatory or assigned (McGrath, 2017). This phenomenon is consistent with the understanding that cognitive dissonance is more complicated than "the animalistic fear" (Akerlof and Dickens, 1982, p. 311) of negative outcomes directly resulting from past actions. It can arise when individuals perceive their past behavior as a reflection of being incompetent or immoral, which undermines their social and self-image (Aronson et al., 1968, Harmon-Jones and Mills, 2019).[21]

To incorporate these elements of cognitive dissonance, I explicitly model the DM's type as a dimension of the state space. In an extension of the CD game, higher-type individuals are more likely to make correct choices in the short run and earn higher payoffs in the long run, consistent with reputation concern models (Prendergast and Stole, 1996). In this framework, belief distortion is more significant in free-choice scenarios than in assigned-choice scenarios, as voluntarily making a poor decision signals incompetence and hurts the DM's self-image.

Setup An investor has $\delta > 0$, $\theta > 0$, and $u(x) = x$. There are two assets Z_1, Z_2, with only one, denoted Z^*, being truly valuable. The state space $\Omega = \{Z_1, Z_2\} \times \{h, l\}$ represents which asset is the valuable one and whether the investor's type is high (h) or low (l). The investor initially believes the two dimensions are independent, attaching probability $\frac{1}{2}$ to Z_1 and $\frac{2}{3}$ to h.

[20] Vaidis et al. (2024) conducted a multi-lab replication of the induced compliance experiment by Festinger and Carlsmith (1959). While the replication supports the key prediction that attitudes shift to align with past actions, it found "no significant difference in attitude" between conditions where participants perceived high versus low freedom of choice.

[21] This idea is at the core of many classical studies on cognitive dissonance. For example, in Festinger and Carlsmith (1959), dissonance arises after individuals are instructed to lie to someone else, creating a conflict with their wish to be seen as honest.

In this setting, the timeline of the CD game is as follows. The investor chooses an asset at $t = 1$, earning a payoff of 0 if her choice is truly valuable and $-L$ otherwise. After this investment, she optimally distorts her beliefs. At $t = 2$, the investor makes no choice; during this period, high types earn a payoff of 1, while low types earn 0 due to their weaker reputation or lower competency.[22]

I extend the CD game by including one initial stage before the game starts, in which the investor is assigned to either of the following two scenarios:

- *free-choice*: the investor knows her period-1 feasible set in the CD game will contain both assets. Before making a choice, she does some costless analysis (thinking) about which asset is valuable, which identifies the correct asset Z^* for high types (h) and mistakenly identifies the other if she is a low type (l). Observe that, after thinking about which asset is valuable, her choice is the truly valuable asset with probability $\frac{2}{3}$.

- *assigned-choice*: the investor knows she will not make a choice in the CD game. Instead, the choice is delegated to a mutual fund manager, who *assigns* her an asset $Z_0 \in \{Z_1, Z_2\}$ by the start of the CD game. Regardless of the investor's type, the assignment Z_0 matches the truly valuable asset with probability $\frac{2}{3}$.

Beyond the financial market interpretation in line with Chang et al. (2016), this setting can also be understood in the context of laboratory studies on cognitive dissonance. In this interpretation, the "investor" represents an experimental participant, while the "mutual fund manager" corresponds to the experimenter. Under this interpretation, the psychology literature typically suggests that participants experience more dissonance in the *free-choice* scenario than in the *assigned-choice* scenario (Harmon-Jones and Mills, 2019).

Without loss of generality, for both scenarios, I assume the investor's period-1 investment is Z_1 and analyze the investor's period-2 beliefs under this assumption.[23]

DEFINITION 4.1. *For $L > 0$, I define $P^*_{Z_1,free}(L)$ as the investor's equilibrium period-2 belief that Z_1 is truly valuable in the free-choice scenario. Similarly, I define $P^*_{Z_1,assign}(L)$ as the investor's equilibrium period-2 belief that Z_1 is truly valuable in the assigned-choice scenario.*

[22] This simple setting can be extended to account for decision-making at $t = 2$. For instance, at $t = 2$, the investor may incur a cost to retract her investment or switch to another asset. I omit this for simplicity in the exposition.

[23] For the *free-choice* scenario, this means the investor identifies asset Z_1 after thinking about it. For the *assigned-choice* scenario, this means the investor is assigned asset Z_2 by the mutual fund manager.

PROPOSITION 4.3. *There exists $\epsilon > 0$, such that for all potential loss from the investment $L > 0$, it holds that*

$$P^*_{Z_1,\,free}(L) > P^*_{Z_1,\,assign}(L) + \epsilon.$$

Proposition 4.3 states that in the free-choice scenario, the investor exhibits a stronger distortion in beliefs that favor the initial investment. In this case, a poor decision would be seen as the investor's *own fault*, negatively affecting her self-image. This creates a stronger incentive to justify her choice as superior to the forgone alternative. On the other hand, in the assigned-choice scenario, the DM's self-perceived competence is independent of which asset is more valuable, thereby attenuating the cognitive dissonance after making the period-1 investment.

In the proposition, the belief gap between free- and assigned-choice scenarios persists even as the loss from making a poor investment choice approaches zero ($L \to 0$). In this case, while the direct negative outcome from the investment disappears, the desire for a positive self-image still leads to belief distortion in favor of Z_1 in the free-choice scenario.[24]

5. APPLICATIONS

In this section, I discuss three applications: how cognitive dissonance may shape buyer-seller interactions, impose a "cognitive tax" on people in poverty, and affect the value of information.

5.1 *Upselling through Upgrading*

I apply the CD game to model upselling, where a monopolist offers add-ons after customers commit to a basic purchase. While cognitive dissonance clearly allows the monopolist to charge a higher price for the add-on, its effect on the equilibrium *quantity* is less clear. I show that the seller optimally employs a *mass marketing* strategy that expands the quantity sold for both the basic product and the add-on. Consumer welfare is always strictly higher compared to a scenario where both products are sold simultaneously, which contrasts with the welfare outcomes found in existing studies of add-on selling (Ellison, 2005, Gabaix and Laibson, 2006). When the dissonance factor θ is sufficiently low, upselling is strictly profitable for the seller.

[24]Here, the utility of a positive self-image comes from the expectation of future reward. One can get similar results by modeling the utility from a positive self-image as stemming from ego utility (Kőszegi, 2006).

To illustrate the role of cognitive dissonance in upselling, consider a gym customer who is unsure whether she will become a "gym person" who enjoys working out (state g) or someone who does not (state b). After signing up for a gym membership, the desire to rationalize her decision to join leads her to overestimate the likelihood of g.[25] This bias makes her more susceptible to upselling strategies; for example, the gym staff can more easily persuade her to purchase an expensive personal training package, which is more appealing to those who believe they are "gym people."

Upselling is a common business practice. Car dealerships frequently offer upgrade options like extended warranties during the final stages of a purchase. The travel and hospitality industry often promotes pricier upgrades by sending reminders after online bookings and making offers at check-in. Digital service providers often start with a low-cost or free basic membership, then subtly encourage customers to upgrade to a premium plan after they join ("freemium-to-premium"). The market recognizes the profitability of upselling. For example, companies like AfterSell[26] specialize in producing mobile apps for online retailers that automatically generate upsell offers after a customer completes an online purchase.

Seller A monopolist ("seller") produces a basic good and an add-on with zero marginal cost. The state space is $\Omega = \{H, L\}$. In state H, the basic good generates payoff 1, and the add-on generates payoff $\alpha \in (0, 1]$; in state L, both goods generate payoff 0.

Consumer There is a unit measure of consumers with type π distributed over $[0, 1]$ according to a full-support distribution with a differentiable CDF F. A consumer with type π believes that $Prob(H) = \pi$ and $Prob(L) = 1 - \pi$ *ex-ante*. All consumers share the same $\theta > 0$, $\delta = 1$ and $u(x) = x$. The inverse demand function for the basic good is given by $\hat{P} = [1 - F(\pi)]^{-1}$, while the inverse demand for the add-on is $\alpha\hat{P}$.

The Upselling Game At time $t = 1$, the seller offers a price $P_1 > 0$ for the basic good to all consumers. The period-1 self of each consumer then chooses between (1) purchasing one unit of the basic good at a price P_1 and (2) an outside option that terminates the game and

[25] Della Vigna and Malmendier (2006) demonstrates that gym members often exhibit overconfidence regarding their future workout frequency, with those opting for a monthly membership effectively paying over \$17 per expected visit – significantly more than the \$10 per visit cost associated with a 10-visit pass. In my example, the mitigation of cognitive dissonance amplifies this overconfidence following the initial purchase of the membership.

[26] The company's product enables online retailers to present customers with an upsell offer that can be accepted with a single click, without requiring re-entry of payment or shipping information. Case studies from AfterSell suggest that their post-purchase upselling tool leads to a significant increase in seller revenues (https://www.aftersell.com/case-studies).

pays zero. For consumers who purchase the basic good, the game proceeds to the next stage, where the seller offers a price $P_2 > 0$ to all consumers. For each consumer, the intermediate player in the CD game then selects a distorted belief, which determines whether the consumer purchases the add-on at a price P_2.

For dissonance factor θ and time $t = 1, 2$, let $P^*_{\theta,t}$, $Q^*_{\theta,t}$, and $\Pi^*_{\theta,t}$ define the equilibrium price, quantity, and revenue for period t respectively. Define the *ex-ante* consumer surplus as

$$CS^*_\theta = \sum_{t\in\mathcal{T}} \int_0^{Q^*_{\theta,t}} [\hat{P}(Q) - P^*_{\theta,t}]dQ.$$

An equilibrium is non-degenerate if $Q^*_{\theta,t} > 0$ for $t = 1, 2$. I assume the existence of non-degenerate equilibria and focus solely on such cases.[27]

The Static Benchmark My analysis compares the equilibrium outcomes of the upselling game with those of the static benchmark, where the seller offers both products together as a bundle at $t = 1$, and makes no additional offer at $t = 1$. The consumers choose whether or not to purchase at $t = 1$. In this scenario, cognitive dissonance plays no role, and the equilibrium prices (*resp.* quantities) for the basic product and the add-on are $P^*_{0,1}$ and $P^*_{0,2}$ (*resp.* $Q^*_{0,1}$ and $Q^*_{0,2}$). These are exactly the equilibrium prices (*resp.* quantities) in the upselling game with $\theta = 0$. I assume the CDF F of the type distribution guarantees the uniqueness of $P^*_{0,t}$ and $Q^*_{0,t}$ for $t = 1, 2$.

PROPOSITION 5.1. *(Mass Marketing) For $\theta > 0$, the equilibrium quantities satisfy*

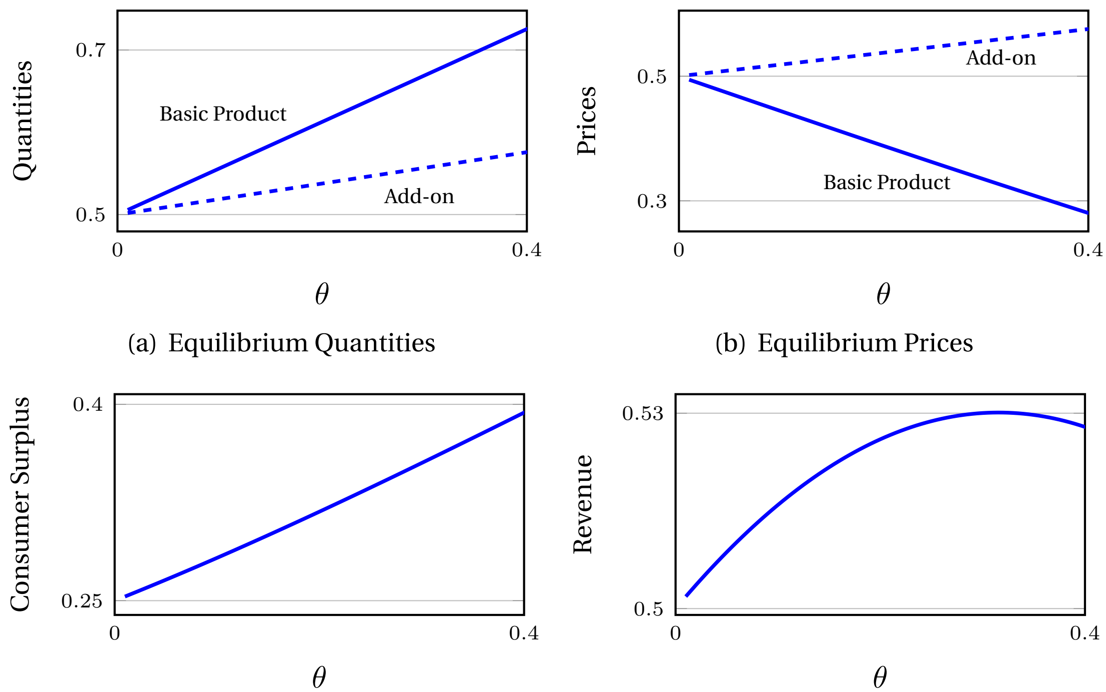
$$Q^*_{\theta,1} > Q^*_{\theta,2} > Q^*_{0,2} = Q^*_{0,1}.$$

Proposition 5.1 states that, compared to the static benchmark, the seller in the upselling game adopts the *mass marketing* strategy, aiming to sell both the basic product and the add-on to a broad customer base rather than pursuing a *niche marketing* approach that targets a more focused consumer segment.

In Figure 5.1, I illustrate the result assuming that the consumer type π is uniformly distributed and that $\alpha = 1$. Part (a) of the figure illustrates the *mass selling* approach: as the

[27] Such an equilibrium exists if, for example, $\theta \in \left(0, \frac{\ln 2}{2}\right)$. This assumption rules out the case where the *ex-post* belief for some type π exceeds 2π, where the evaluation of the product increases by over 100% due to dissonance mitigation. It rules out the market collapse that could arise when the magnitude of belief distortion is extreme, where the seller would charge an extremely high price for the second period, which keeps the forward-looking consumers away from the market *ex-ante*.

dissonance factor θ increases, the equilibrium quantities sold for both the basic product and the add-on increase. Parts (b)-(d) illustrate the equilibrium prices, consumer surplus, and seller's revenue, which will be formally discussed in Propositions 5.2 through 5.4 below.



(a) Equilibrium Quantities

(b) Equilibrium Prices

(c) Consumer Surplus

(d) Revenue

FIGURE 5. Equilibrium outcomes as functions of the dissonance factor θ

As illustrated in Figure 5.1 (a), for the basic product, $Q^*_{\theta,1}$ exceeds benchmark quantity $Q^*_{0,1}$ – the seller leverages cognitive dissonance by encouraging more customers to purchase the basic product, thereby increasing $Q^*_{\theta,1}$. A less straightforward aspect of my result is that the seller continues mass marketing in the second period, expanding add-on sales to less optimistic customers (those with lower π). This is profitable because these *less optimistic* consumers are *more susceptible* to the effects of cognitive dissonance in enhancing confidence in the product's quality. For example, in the extreme case where a consumer has $\pi = 1$ and hence is fully confident in state H, cognitive dissonance cannot increase her confidence further. On the other hand, the smaller π is, the greater the potential for cognitive dissonance to enhance the willingness to purchase the add-on.[28]

[28] In the analysis above, I explain that increasing add-on sales quantity benefits the seller by exploiting higher profits from cognitive dissonance. However, there is a trade-off – since consumers are forward-looking, increasing add-on sales requires lowering the total price of the basic product and add-on, which reduces the profit at

PROPOSITION 5.2. *(Foot-in-the-Door) If $\theta > 0$ and π is uniformly distributed, then*

$$P^*_{\theta,1} < P^*_{0,1} = \frac{1}{\alpha} P^*_{0,2} < \frac{1}{\alpha} P^*_{\theta,2}.$$

As illustrated by Figure 5.1 (b), Proposition 5.2 compares the basic product price $P^*_{\theta,1}$ and the normalized add-on price $\frac{1}{\alpha} P^*_{\theta,2}$ with the rational benchmark. In equilibrium, the seller reduces the basic product price while charging a premium for the add-on to exploit the cognitive biases of existing customers. The pricing strategy exemplifies the *foot-in-the-door* selling technique: obtaining a customer's initial agreement to purchase the basic product facilitates charging a higher price for the add-on later.

PROPOSITION 5.3. *(Enhanced Consumer Welfare) If $\theta > 0$, then*

$$CS^*_\theta > CS^*_0.$$

In contrast to the welfare implications of existing models of add-on selling (Ellison, 2005, Gabaix and Laibson, 2006),[29] consumers *benefit* from the seller's upselling strategy. In Figure 5.1 (c), the *ex-ante* customer surplus increases with the cognitive dissonance factor θ. To see how this is possible, notice that cognitive dissonance allows the seller to adopt a mass marketing strategy that "enlarges the pie," increasing the total surplus in the market. This enables the slice belonging to customers to grow larger.

The mass marketing strategy benefits the consumers both *directly* and *indirectly*. First, as Proposition 5.1 shows, a positive mass of consumers benefits directly: they would be excluded in the static benchmark but purchase a product in the upselling game. Second, mass marketing also creates a *positive externality* for other consumers: since consumers are forward-looking, the seller can only increase the volume of add-on sales by lowering the combined price of the basic product and add-on, $P^*_{\theta,1} + P^*_{\theta,2}$. [30] This benefits all consumers.

$t = 1$. Mathematical analysis shows that this effect does not dominate in determining the equilibrium quantity of add-on sales.

[29]The driving force for add-on selling in my paper is distinct from both papers. In Ellison (2005), discriminatory pricing for add-ons helps the seller soften competition. In Gabaix and Laibson (2006), add-on pricing exploits naive customers. Unlike both papers, I do not assume that add-on prices remain concealed until the point of sale. My assumption appears plausible for add-on sales conducted by online platforms, where customers can easily access information about prices.

[30]If customers are naive, the seller might initially set a higher total price than the one-time benchmark. The seller would be compelled to reduce the price if naive customers gradually learn to anticipate the seller's upselling tactics.

PROPOSITION 5.4. *(Profitability of Upselling) If $\theta > 0$ and π is uniformly distributed, then there exists a dissonance factor $\bar{\theta} > 0$, such that for all $\theta \in (0, \bar{\theta})$,*

$$\sum_{t\in\mathcal{T}} \Pi^*_{\theta,t} > \sum_{t\in\mathcal{T}} \Pi^*_{0,t}.$$

Proposition 5.4 states that using the upselling strategy strictly increases the seller's revenue if the dissonance factor θ is sufficiently small. However, as is illustrated in Figure 5.1 (d), this result may not hold for larger values of θ. When θ is very large, the seller will charge a high add-on price, as she cannot commit to the add-on price in the first period. This makes attracting forward-looking customers in the first period costly, ultimately reducing total revenue.

5.2 *Mitigating Poverty-Related Dissonance*

In this section, I discuss how poverty and cognitive dissonance can mutually reinforce each other. Individuals with *lower* baseline incomes may experience *greater* dissonance from their past actions.[31] Holding the dissonance factor fixed, this heightened dissonance reduces the quality of decision-making for those in poverty, effectively imposing a "cognitive tax" that exacerbates the poverty burden. To pin down the idea, consider the following example.

EXAMPLE 2. Alice and Bob are individuals who lost their ability to work due to a chronic illness. Alice receives a $250 monthly subsistence allowance, while Bob gets $25. Initially, without modern healthcare, Alice and Bob each spent $20 each month to see a traditional healer or herbalist. Recently, a modern hospital has opened in their village. Both agents receive brochures urging them to seek modern treatment, which also costs $20 per month. These brochures, backed by expert recommendations, also emphasize that modern treatment is more effective than the alternative therapies they previously used.

If both agents trust the brochure, Bob is likely to feel *more* cognitive dissonance than Alice. For Alice, the $20 spent on alternative treatments is just 8% of her income, making it easier to accept the treatments' ineffectiveness. For Bob, it's 80% of his income, a significant

[31] In the real world, poverty and low social status often leads to anxiety (Ridley et al., 2020, Sergeyev et al., 2023) and cognitive dissonance (Oxoby, 2003, 2004), especially between the desire for higher status and the fear of staying in poverty. To reduce this dissonance, disadvantaged individuals may downplay the value of status-seeking investments like education or career advancement (Penn, 2017). Similarly, disadvantaged groups may turn to religious beliefs, which offer "illusory happiness" by promising afterlife rewards, compensating for the possibility that real-life efforts may not lead to worldly success (Marx, 1844, Chen, 2010).

sacrifice. This makes him more likely to reject the brochure and stick with the alternative treatments, which justifies his past decisions and avoids the feeling of wasted sacrifices. □

Next, I introduce an application of the CD game that formalizes the idea above.

Setup Consider a DM with $\theta > 0$, $\delta \geq 0$ and a strictly concave utility index. There are two periods, $t = 1, 2$, and two technologies, o (old) and ν (new). In the context of Example 2, o represents the alternative treatment, and ν is the modern one. The state space is $\Omega = \{o, \nu\}$. The modeler knows that the true state $\omega^* = \nu$, that is, the new technology is truly productive, but the DM only knows ω^* remains unchanged over time, with an *ex-ante* belief $\mu_1 \in \Delta\Omega$ with full support.

For each period, the baseline income in each period is given by $Y > 0$. For baseline income Y and technology choice o, the corresponding consumption is defined by

$$c^{Y,o}_{\omega} = \begin{cases} Y + 1 & \text{if } \omega \text{ equals } o \\ Y & \text{if } \omega \text{ equals } \nu. \end{cases}$$

Similarly, for technology choice ν, the corresponding consumption $c^{Y,\nu}$ yields payoff $Y + 1$ if $\omega^* = \nu$ and Y otherwise. The DM is endowed with period-1 action $(c^{Y,o}, \{c^{Y,o}, c^{Y,\nu}\})$. That is, at $t = 1$, the DM has to choose the old technology o, representing a prohibitively high cost to use ν. At $t = 2$, the DM may choose to switch from o to ν.

I consider an extension of the CD game where, during the interim stage between $t = 1$ and $t = 2$, the DM receives a signal realization $s \in S = [0, 1]$.[32] The signal's probability density function (*pdf*) is $f(s) = 2s$ if $\omega^* = o$ and $f(s) = 2 - 2s$ if $\omega^* = \nu$. For $s \in S$, denote the Bayesian posterior as μ_1^s.[33] The DM distorts μ_1^s according to

$$\max_{\mu \in \Delta\Omega} \theta U_{\mu}(c^{Y,o}, c_2^*) - D_{KL}(\mu || \mu_1^s), \tag{12}$$

with the correct anticipation on how distorting the Bayesian posterior influences the period-2 consumption choice c_2^*. At $t = 2$, the DM has the distorted belief and chooses $c_2^* \in \{c^{Y,o}, c^{Y,\nu}\}$ that maximizes her expected utility. For a given dissonance factor θ and baseline

[32]This type of extension, which allows for belief updating in the interim stage, will be formally introduced in Section 5.3.

[33]My results continue to hold if this information structure is replaced by any other unbounded information structure with atomless signal distributions.

income Y, let $P^*(\nu \mid \theta, Y)$ represent the probability that the DM adopts the new technology (ν) at $t = 2$.

PROPOSITION 5.5. *For $\theta, Y > 0$, $P^*(\nu \mid \theta, Y)$ is strictly decreasing in θ and strictly increasing in Y.*

In Proposition 5.5, people with greater dissonance factor θ are more reluctant to adopt the new technology due to the desire to justify their previous investment in o. More notably, individuals with lower baseline incomes are even more reluctant to switch. This reluctance stems from a heightened fear of failure in the initial technology choice among those with lower baseline income. Formally, the utility loss from a failed period-1 investment, $u(Y+1) - u(Y)$, grows as Y declines due to diminishing marginal utility. Consequently, as Y decreases, belief distortion intensifies in the interim period, reducing the likelihood of technology adoption.

Proposition 5.5 shows that cognitive dissonance can create inertia in technology adoption among poorer individuals. In Appendix C.2, I provide an additional result showing that cognitive dissonance can also lead to unnecessary risk-taking, as the need to justify the initial investment in o pushes poorer individuals to over-invest in the old technology at $t = 2$. This distinguishes my framework from alternative explanations, such as higher risk aversion or tighter credit constraints among poor households.

I define the excess income as the difference between the DM's total income and baseline income. Since $\omega^* = o$, the excess income for the second period equals 1 if the DM switches to ν and 0 if she continues with o. Therefore, the expected excess income equals the probability of technology adoption, $P^*(\nu \mid \theta, Y)$. The following corollary then follows.

COROLLARY 5.1. *For $\theta > 0$, the expected excess income at $t = 2$ is strictly increasing in Y.*

In Corollary 5.1, increasing the baseline income Y strictly raises the expected excess income. This is because earning a higher baseline income mitigates cognitive dissonance, which enhances the quality of decision-making in period 2. Conversely, poverty imposes a "cognitive tax" by amplifying cognitive dissonance.

5.3 *Value of Information*

In this section, I extend the CD game by allowing the DM to update her beliefs based on new information received during the interim stage. With this extended framework, I study the *ex-ante* value of information and show the following two results. First, beyond the common wisdom that dynamically inconsistent agents may use information avoidance as a commitment device (Carrillo and Mariotti, 2000, Golman et al., 2017), high-quality information reduces the dissonance associated with uncertainties in payoffs. Hence, such information can be more valuable to individuals prone to cognitive dissonance than the rational benchmark. Second, as the dissonance factor θ grows, the range of prior beliefs that can be negatively impacted by information *diminishes*, contrary to the intuition that stronger cognitive biases make information avoidance more appealing as a self-commitment device.

I consider an extension of the CD game that allows belief updating. Fix $u \in \mathcal{U}$ and $\delta \in (0,1)$. I assume that the DM is initially endowed with a singleton menu $A = \{(c_1, C_2)\}$.[34] In the interim stage, the DM updates her belief according to a Blackwell experiment ρ, which is formalized by a distribution over Bayesian posteriors, that is, over $\Delta\Omega$. Let $\mu(\rho) = \mathbb{E}_\rho[\mu]$ denote the prior belief corresponding to ρ and let $\mathcal{B}$ denote the set of all experiments. In the interim stage, Nature selects a realized posterior $\tilde{\rho}$ according to the experiment ρ, and then the intermediate player takes $\tilde{\rho}$ as the reference point and selects a belief $\mu^{\tilde{\rho}}$ that maximizes

$$\theta U_\mu(c_1, c_2^*(\mu)) - D_{KL}(\mu || \tilde{\rho}).$$

Here, $c_2^*(\mu)$ defines Player 2's optimal choice of period-2 consumption given the belief choice μ. Finally, Player 2 chooses $c_2^*(\mu^{\tilde{\rho}}) \in C_2$ to maximize the expected utility $\mathbb{E}_{\mu^{\tilde{\rho}}}[u(c_2)]$. The *equilibrium consumption stream* is given by $\boldsymbol{c}^* = \left[c_1, (c_2^*(\mu^{\tilde{\rho}}))_{\tilde{\rho} \in Supp(\rho)}\right]$, with the corresponding *ex-ante* expected utility be given by

$$\mathbb{U}_\rho(\boldsymbol{c}^*) = \mathbb{E}_{\mu(\rho)}[u(c_1)] + \delta \int_{\Delta\Omega} \mathbb{E}_{\tilde{\rho}}[u(c_2^*(\mu^{\tilde{\rho}}))] d\rho. \tag{13}$$

For period-1 action $a = (c_1, C_2)$, experiment $\rho \in \mathcal{B}$, and dissonance factor $\theta \geq 0$, let $SPNE_{\theta,\rho}(\{a\})$ denote the set of all equilibrium consumption streams. The *ex-ante* value of experiment ρ is

$$V_a(\rho, \theta) = \sup_{\boldsymbol{c} \in SPNE_{\theta,\rho}(\{a\})} \mathbb{U}_\rho(\boldsymbol{c}) - \sup_{c \in SPNE_\kappa(\{a\})} U_{\mu(\rho)}(c),$$

where $\kappa = (\mu(\rho), u, \delta, \theta)$ is the parameter for the CD game where the DM has prior $\mu(\rho)$ and no access to information. In this expression, I take the supremum to reflect the equilibrium

[34] All definitions in this section can be generalized to cases with larger menus.

selection criterion that favors the *ex-ante* self. The term $V_a(\rho,\theta)$ measures the value of information ρ for a DM with dissonance factor θ, which is the difference between the expected utility when the DM learns from experiment ρ ($\mathbb{U}_\rho(\boldsymbol{c})$) and the expected utility when no information arrives ($U_{\mu(\rho)}(c)$).

A period-1 action $a=(c_1,C_2)$ is *regular* if C_2 is finite and $V_a(\rho_0,\theta_0)\neq V_a(\rho_0,0)$ for some $\theta_0>0$ and $\rho_0\in\mathcal{B}$. In other words, cognitive dissonance may influence the value of information for at least one Blackwell experiment.

PROPOSITION 5.6. *For every regular period-1 action a, there exists an interval $(0,\hat{\theta})$ such that*

- *for all $\theta\in(0,\hat{\theta})$, there exists $\rho\in\mathcal{B}$ such that*

$$V_a(\rho,\theta)<0\leq V_a(\rho,0); \tag{14}$$

- *for all $\theta>0$, there exists $\rho'\in\mathcal{B}$ such that*

$$V_a(\rho',\theta)>V_a(\rho',0)>0. \tag{15}$$

For a DM with dissonance factor $\theta\in(0,\hat{\theta})$, Proposition 5.6 establishes the coexistence of *information avoidance* (as shown in (14)) and a *stronger preference for information acquisition* (as shown in (15)). There exists an experiment ρ that yields a negative value, while another experiment ρ' offers a higher value than in the rational benchmark.

In the proposition, notice that for dissonance factor $\theta\in[\hat{\theta},\infty)$, the DM might not exhibit information avoidance. Next, I will formalize that *information avoidance* can become less frequent as the dissonance factor θ increases.

DEFINITION 5.1. *For dissonance factor θ, we say information can hurt prior belief μ^* if there exists a ρ such that $\mu(\rho)=\mu^*$ and $V_a(\rho,\theta)<0$. Let $P_{Neg}(a,\theta)$ denote the collection of all priors that can be hurt by information.*

The question I aim to address is how $P_{Neg}(a,\theta)$, the range of prior beliefs that can be negatively impacted by information, varies with the dissonance factor θ. I demonstrate that the size of $P_{Neg}(a,\theta)$ can decrease as θ increases.

PROPOSITION 5.7. *Suppose* $|\Omega| = 2$. *For every regular* a, *there exists* $\theta^* > 0$ *such that for all* $\bar{\theta} > \underline{\theta} > \theta^*$, *it holds that*

$$P_{Neg}(a, \bar{\theta}) \subset P_{Neg}(a, \underline{\theta});$$

moreover, if both sets are non-empty, then $P_{Neg}(a, \bar{\theta}) \subsetneq P_{Neg}(a, \underline{\theta})$.

Proposition 5.7 states that for a sufficiently large dissonance factor θ, the larger the θ, the less likely it is that new information will harm prior beliefs. In general, the size of priors that can be hurt by information exhibits a non-monotonic relationship with respect to θ. When $\theta = 0$, information is always beneficial and $P_{Neg} = \emptyset$. As θ increases, the region P_{Neg} initially expands, meaning information can harm a broader range of prior beliefs, and then decreases as $\theta \to \infty$.

To build intuition for Proposition 5.7, we consider two central questions: (1) What drives information avoidance among individuals who perceive cognitive dissonance? (2) Why does information avoidance diminish as the dissonance factor gets larger?

To answer these questions, let's consider a simple example where $\Omega = \{L, R\}$, $a = (c_1, C_2)$, with $c_1 = (1, 1)$ providing state-independent payoffs, and $C_2 = \{(-2, 2), (1, 1), (2, -2)\}$ (all payoffs in utils). In this case, it can be shown that $P_{\text{Neg}}(a, \theta) = (\epsilon_\theta, 1 - \epsilon_\theta)$ for some $\epsilon_\theta \in (0, \frac{1}{2}]$.[35] This implies that information may only harm individuals with *weaker* priors (those less certain) who choose the riskless option $(1, 1)$ when no information is available. When some information is accessible, learning about the true state can lead the DM to find riskier options more appealing. In such cases, dissonance mitigation can distort decision-making, "luring" the DM into excessive risk-taking at $t = 2$ (Akerlof and Dickens, 1982). If this distortion becomes severe enough, information may indeed harm the DM.

To understand why information avoidance diminishes as dissonance increases, note that as $\theta \to \infty$, the drive to mitigate dissonance becomes overwhelming. With most prior beliefs, the DM would choose risky options $(-2, 2)$ and $(2, -2)$ even without additional information. Thus, the mechanism behind information avoidance—where updating leads to excessive risk-taking—no longer applies. In this example, $P_{\text{Neg}}(a, \theta)$ is empty for all sufficiently large θ, indicating that information never harms when dissonance mitigation becomes dominant.

[35] In Appendix C.3, I formally present the proposition for this observation in a more general setting, where the period-2 menu C_2 can be of any size.

REMARK. In general, a DM susceptible to cognitive dissonance has a stronger preference for high-quality information (which brings the posterior close to certainty) than the rational benchmark. Beyond the instrumental value, learning such information also makes the DM nearly convinced she knows the actual state, which *reduces* cognitive dissonance related to payoff uncertainties and improves the quality of decision-making. Proposition C.3 in Appendix C.3 formalizes this intuition.

6. DISCUSSION

In this section, I discuss the CD game's assumptions, including Player 1's sophistication, the intermediate player's forward-looking behavior, and the choice of the cost function for belief distortion.

6.1 *Sophistication*

Player 1 and the intermediate player are sophisticated – they have rational expectations regarding how mitigating cognitive dissonance distorts future choices. There are two reasons for assuming sophistication. First, for the intermediate player, it is less plausible that she would choose beliefs for her future self without recognizing that these choices will influence her future decision-making. Second, for Player 1, relaxing dynamic consistency while maintaining sophistication minimizes the deviation from the rational benchmark. As demonstrated in Sections 4 and 5, even without assuming *naïveté*, cognitive dissonance leads to various implications that the standard model cannot account for. Assuming sophistication also allows the discussion of the optimal preventive measures the agent may take to mitigate the distortions created by cognitive dissonance.

6.2 *Forward-Looking intermediate player*

The intermediate player may experience cognitive dissonance after Player 1 commits to a set of future options, formally represented by the menu C_2 in the period-1 action (c_1, C_2). This assumption is consistent with laboratory evidence showing that committing to a future action can induce cognitive dissonance even before the action is taken (Beauvois et al., 1995). This assumption is also natural due to the forward-looking nature of the DM – by the interim stage, the payoff from both c_1 and the equilibrium $c_2^* \in C_2$ have not yet been realized. If the intermediate player experiences dissonance from negative outcomes tied to c_1, it is logical that she would experience dissonance from those associated with c_2^* as well.

Assuming that the intermediate player is forward-looking is also useful for the model's applications. For instance, in Section 4.3, the intermediate player distorts her beliefs more strongly in anticipation that higher types will earn greater payoffs in the long run, explaining why belief distortion is greater in the free-choice scenario. In Section 5.3, the forward-looking nature of the intermediate player allows an analysis of how cognitive dissonance influences the value of information.

6.3 *Cost of Belief Distortion*

Following Hansen and Sargent (2001) and Caplin and Leahy (2019), I take the KL divergence as the cost function for belief distortion, which provides a parsimonious yet powerful framework for analyzing cognitive dissonance. Of all divergence measures taking the form $D(\mu||\mu_1)$, the KL divergence is the *unique* one that rules out violations of Savage's sure-thing principle (Strzalecki, 2011), thus preventing Ellsberg-type ambiguity aversion (or seeking) from interfering with the analysis of cognitive dissonance. An alternative cost function could be the difference between the value of the equilibrium consumption stream and the *ex-ante* optimal consumption stream. In general, the costs calculated using these two measures are positively correlated. For instance, when $|\Omega| = 2$, both cost functions are U-shaped, with the minimum achieved at μ_1.

References

Acharya, Avidit, Matthew Blackwell, and Maya Sen (2018), "Explaining preferences from behavior: A cognitive dissonance approach." *The Journal of Politics*, 80(2), 400–411. [6, 11, 24, 84]

Aczél, János (2005), "Extension of a generalized pexider equation." *Proceedings of the American Mathematical Society*, 133(11), 3227–3233. [62]

Akerlof, George A and William T Dickens (1982), "The economic consequences of cognitive dissonance." *The American economic review*, 72(3), 307–319. [2, 5, 6, 7, 8, 11, 12, 18, 28, 40, 84]

Anscombe, Francis J and Robert J Aumann (1963), "A definition of subjective probability." *Annals of mathematical statistics*, 34(1), 199–205. [12, 14]

Aronson, Elliot and Judson Mills (1959), "The effect of severity of initiation on liking for a group." *The Journal of Abnormal and Social Psychology*, 59(2), 177. [6, 23, 27]

Aronson, Elliot et al. (1968), "Dissonance theory: Progress and problems." *Theories of cognitive consistency: A sourcebook*, 249, 253. [12, 28]

Beauvois, Jean-Léon, Marc Bungert, and Pascale Mariette (1995), "Forced compliance: Commitment to compliance and commitment to activity." *European journal of social psychology*, 25(1), 17–26. [6, 41]

Bénabou, Roland and Jean Tirole (2004), "Willpower and personal rules." *Journal of Political Economy*, 112(4), 848–886. [8]

Benabou, Roland and Jean Tirole (2006), "Belief in a just world and redistributive politics." *The Quarterly journal of economics*, 121(2), 699–746. [8, 23]

Bénabou, Roland and Jean Tirole (2011), "Identity, morals, and taboos: Beliefs as assets." *The Quarterly Journal of Economics*, 126(2), 805–855. [6, 8]

Bernheim, B Douglas, Luca Braghieri, Alejandro Martínez-Marquina, and David Zuckerman (2021), "A theory of chosen preferences." *American Economic Review*, 111(2), 720–754. [7]

Bertrand, Marianne, Sendhil Mullainathan, and Eldar Shafir (2004), "A behavioral-economics view of poverty." *American Economic Review*, 94(2), 419–423. [2]

Bracha, Anat and Donald J Brown (2012), "Affective decision making: A theory of optimism bias." *Games and Economic Behavior*, 75(1), 67–80. [7]

Brehm, Jack W (1956), "Postdecision changes in the desirability of alternatives." *The Journal of Abnormal and Social Psychology*, 52(3), 384. [6, 24]

Brunnermeier, Markus K and Jonathan A Parker (2005), "Optimal expectations." *American Economic Review*, 95(4), 1092–1118. [7]

Burgh, Jarrod and Emerson Melo (2024), "Wishful thinking is risky thinking." *Available at SSRN 4746142*. [7]

Caplin, Andrew and John Leahy (2001), "Psychological expected utility theory and anticipatory feelings." *The Quarterly Journal of Economics*, 116(1), 55–79. [5, 7]

Caplin, Andrew and John V Leahy (2019), "Wishful thinking." Technical report, National Bureau of Economic Research. [3, 7, 42]

Carrillo, Juan D and Thomas Mariotti (2000), "Strategic ignorance as a self-disciplining device." *The Review of Economic Studies*, 67(3), 529–544. [6, 7, 38]

Chambers, Christopher P, Yusufcan Masatlioglu, and Collin Raymond (2023), "Coherent distorted beliefs." *arXiv preprint arXiv:2310.09879*. [7]

Chang, Tom Y, David H Solomon, and Mark M Westerfield (2016), "Looking for someone to blame: Delegation, cognitive dissonance, and the disposition effect." *The Journal of Finance*, 71(1), 267–302. [2, 4, 6, 11, 12, 23, 27, 28, 29, 84]

Chauvin, Kyle P (2021), "Belief updating with dissonance reduction." Technical report, Working Paper. [12]

Chen, Daniel L (2010), "Club goods and group identity: Evidence from islamic resurgence during the indonesian financial crisis." *Journal of political Economy*, 118(2), 300–354. [35]

Coffman, Katherine Baldiga (2014), "Evidence on self-stereotyping and the contribution of ideas." *The Quarterly Journal of Economics*, 129(4), 1625–1660. [23]

Constantinides, George M (1990), "Habit formation: A resolution of the equity premium puzzle." *Journal of political Economy*, 98(3), 519–543. [21]

Cooper, Joel and Russell H Fazio (1984), "A new look at dissonance theory." In *Advances in experimental social psychology*, volume 17, 229–266, Elsevier. [6, 11]

David, Eric John Ramos (2013), *Internalized oppression: The psychology of marginalized groups*. Springer Publishing Company. [23]

Debreu, Gerard (1959), "Topological methods in cardinal utility theory." [57]

Dekel, Eddie and Barton L Lipman (2012), "Costly self-control and random self-indulgence." *Econometrica*, 80(3), 1271–1302. [7]

Della Vigna, Stefano and Ulrike Malmendier (2006), "Paying not to go to the gym." *american economic Review*, 96(3), 694–719. [31]

Dodoo, Joana Eva and Hosam Al-Samarraie (2021), "A systematic review of factors leading to occupational injuries and fatalities." *Journal of Public Health*, 1–15. [5]

Dupuis, Paul and Richard S Ellis (1997), "A weak convergence approach to the theory of large deviations." *Wiley Series in Probability and Statistics*. [50]

Ellison, Glenn (2005), "A model of add-on pricing." *The Quarterly Journal of Economics*, 120(2), 585–637. [2, 5, 30, 34]

Epstein, Larry G and Igor Kopylov (2007), "Cold feet." *Theoretical Economics*, 2, 231–259. [8]

Eyster, Erik, Shengwu Li, and Sarah Ridout (2022), "A theory of ex post rationalization." *arXiv preprint arXiv:2107.07491*. [5, 7, 23]

Fan, Tony Q (2024), "Forming misspecified mental models: The power of choice." *Available at SSRN 4497643*. [6, 11, 84]

Festinger, Leon (1957), *A theory of cognitive dissonance*. Stanford University Press. [2, 11, 23, 24]

Festinger, Leon and James M Carlsmith (1959), "Cognitive consequences of forced compliance." *The journal of abnormal and social psychology*, 58(2), 203. [4, 6, 8, 11, 24, 28]

Gabaix, Xavier and David Laibson (2006), "Shrouded attributes, consumer myopia, and information suppression in competitive markets." *The Quarterly Journal of Economics*, 121(2), 505–540. [2, 5, 30, 34]

Ghirardato, Paolo (2002), "Revisiting savage in a conditional world." *Economic theory*, 20, 83–92. [19]

Gilboa, Itzhak and David Schmeidler (1989), "Maxmin expected utility with non-unique prior." *Journal of mathematical economics*, 18(2), 141–153. [20, 59]

Gneezy, Ayelet, Alex Imas, Amber Brown, Leif D Nelson, and Michael I Norton (2012), "Paying to be nice: Consistency and costly prosocial behavior." *Management Science*, 58(1), 179–187. [2, 6, 23, 27]

Golman, Russell, David Hagmann, and George Loewenstein (2017), "Information avoidance." *Journal of economic literature*, 55(1), 96–135. [2, 6, 38]

Haagsma, Rein and Niek Koning (2002), "Endogenous mobility-reducing norms." *Journal of Economic Behavior & Organization*, 49(4), 523–547. [6]

Hansen, Lars Peter and Thomas J Sargent (2001), "Robust control and model uncertainty." *American Economic Review*, 91(2), 60–66. [3, 42]

Harmon-Jones, Eddie and Judson Mills (2019), "An introduction to cognitive dissonance theory and an overview of current perspectives on the theory." [4, 12, 22, 23, 24, 28, 29]

Harmon-Jones, Eddie Ed (2019), *Cognitive dissonance: Reexamining a pivotal theory in psychology*. American Psychological Association. [12]

Harris, Christopher (1985), "Existence and characterization of perfect equilibrium in games of perfect information." *Econometrica: Journal of the Econometric Society*, 613–628. [10]

Jakobsen, Alexander M (2021), "An axiomatic model of persuasion." *Econometrica*, 89(5), 2081–2116. [7, 14]

Jost, John T (1997), "An experimental replication of the depressed-entitlement effect among women." *Psychology of Women Quarterly*, 21(3), 387–393. [23]

Jost, John T and Mahzarin R Banaji (1994), "The role of stereotyping in system-justification and the production of false consciousness." *British journal of social psychology*, 33(1), 1–27. [23]

Knox, Robert E and James A Inkster (1968), "Postdecision dissonance at post time." *Journal of personality and social psychology*, 8(4p1), 319. [6, 23, 27]

Konow, James (2000), "Fair shares: Accountability and cognitive dissonance in allocation decisions." *American economic review*, 90(4), 1072–1092. [6, 11, 84]

Köszegi, Botond (2006), "Ego utility, overconfidence, and task choice." *Journal of the European Economic Association*, 4(4), 673–707. [7, 30]

Kovach, Matthew (2020), "Twisting the truth: Foundations of wishful thinking." *Theoretical Economics*, 15(3), 989–1022. [7]

Kremer, Michael, Gautam Rao, and Frank Schilbach (2019), "Behavioral development economics." In *Handbook of behavioral economics: applications and foundations 1*, volume 2, 345–458, Elsevier. [5]

Laibson, David (1997), "Golden eggs and hyperbolic discounting." *The Quarterly Journal of Economics*, 112(2), 443–478. [7]

Lanzani, Giacomo (2022), "Dynamic concern for misspecification." *Available at SSRN 4454504*. [8]

Loewenstein, George (1987), "Anticipation and the valuation of delayed consumption." *The Economic Journal*, 97(387), 666–684. [7]

Maccheroni, Fabio, Massimo Marinacci, and Aldo Rustichini (2006), "Ambiguity aversion, robustness, and the variational representation of preferences." *Econometrica*, 74(6), 1447–1498. [7, 20]

Machina, Mark J (1989), "Dynamic consistency and non-expected utility models of choice under uncertainty." *Journal of Economic literature*, 27(4), 1622–1668. [16]

Mani, Anandi, Sendhil Mullainathan, Eldar Shafir, and Jiaying Zhao (2013), "Poverty impedes cognitive function." *science*, 341(6149), 976–980. [2]

Marx, Karl (1844), "Introduction to a contribution to the critique of hegel's philosophy of right." *Deutsch-Französische Jahrbücher*, 7, 261–71. [35]

Mayraz, Guy (2019), "Priors and desires: A bayesian model of wishful thinking and cognitive dissonance." *Unpublished Manuscript. University of Sydney. http://mayraz.com/papers/PriorsAndDesires. pdf*. [7]

McGrath, April (2017), "Dealing with dissonance: A review of cognitive dissonance reduction." *Social and Personality Psychology Compass*, 11(12), e12362. [28]

Meier, Stephan (2007), "Do subsidies increase charitable giving in the long run? matching donations in a field experiment." *Journal of the European Economic Association*, 5(6), 1203–1222. [24]

Mullainathan, Sendhil and Ebonya Washington (2009), "Sticking with your vote: Cognitive dissonance and political attitudes." *American Economic Journal: Applied Economics*, 1(1), 86–111. [6, 11, 24, 84]

Orhun, Yeşim, Alain Cohn, and Collin B Raymond (2024), "Motivated optimism and workplace risk." *The Economic Journal*, 134(663), 2951–2981. [6]

Oxoby, Robert J (2003), "Attitudes and allocations: status, cognitive dissonance, and the manipulation of attitudes." *Journal of economic behavior & organization*, 52(3), 365–385. [6, 11, 35]

Oxoby, Robert J (2004), "Cognitive dissonance, status and growth of the underclass." *The Economic Journal*, 114(498), 727–749. [6, 35, 84]

Peleg, Bezalel and Menahem E Yaari (1973), "On the existence of a consistent course of action when tastes are changing." *The Review of Economic Studies*, 40(3), 391–401. [7]

Penn, Elizabeth Maggie (2017), "Inequality, social context, and value divergence." *The Journal of Politics*, 79(1), 153–165. [6, 11, 35, 84]

Prendergast, Canice and Lars Stole (1996), "Impetuous youngsters and jaded old-timers: Acquiring a reputation for learning." *Journal of political Economy*, 104(6), 1105–1134. [28]

Rabin, Matthew (1994), "Cognitive dissonance and social change." *Journal of Economic Behavior & Organization*, 23(2), 177–194. [6, 11, 84]

Ridley, Matthew, Gautam Rao, Frank Schilbach, and Vikram Patel (2020), "Poverty, depression, and anxiety: Causal evidence and mechanisms." *Science*, 370(6522), eaay0214. [35]

Sergeyev, Dmitriy, Chen Lian, and Yuriy Gorodnichenko (2023), "The economics of financial stress." Technical report, National Bureau of Economic Research. [35]

Shen, Xiaoxiao (2022), "The effects of forced versus selective exposure to propaganda in china." *Available at SSRN 4177260*. [23]

Sinander, Ludvig (2023), "Optimism, overconfidence, and moral hazard." *arXiv preprint arXiv:2304.08343*. [7]

Siniscalchi, Marciano (2011), "Dynamic choice under ambiguity." *Theoretical Economics*, 6(3), 379–421. [7]

Staw, Barry M (1981), "The escalation of commitment to a course of action." *Academy of management Review*, 6(4), 577–587. [2, 23]

Strotz, RH (1955), "Myopia and inconsistency in dynamic utility maximization." *Review of Economic Studies*, 23(3), 165–180. [7, 16]

Strzalecki, Tomasz (2011), "Axiomatic foundations of multiplier preferences." *Econometrica*, 79(1), 47–73. [3, 42, 55, 58, 59]

Suzuki, Toru (2019), "Choice set dependent performance and post-decision dissonance." *Journal of Economic Behavior & Organization*, 163, 24–42. [6]

Vaidis, David C, Willem WA Sleegers, Florian Van Leeuwen, Kenneth G DeMarree, Bjørn Sætrevik, Robert M Ross, Kathleen Schmidt, John Protzko, Coby Morvinski, Omid Ghasemi, et al. (2024), "A multilab replication of the induced-compliance paradigm of cognitive dissonance." *Advances in Methods and Practices in Psychological Science*, 7(1), 25152459231213375. [4, 6, 8, 11, 23, 28]

Van Veen, Vincent, Marie K Krug, Jonathan W Schooler, and Cameron S Carter (2009), "Neural activity predicts attitude change in cognitive dissonance." *Nature neuroscience*, 12(11), 1469–1474. [6, 11, 23, 24]

Wang, Zichang (2022), "Regret and information avoidance." *Available at SSRN 4269560*. [12]

Werner, Jan (2005), "A simple axiomatization of risk-averse expected utility." *Economics Letters*, 88(1), 73–77. [58]

Yang, Tony Zirui (2022), "Participatory censorship in authoritarian regimes." *Available at SSRN 4167554*. [23]

Yariv, Leeat (2005), "I'll see it when i believe it." Technical report, mimeo. [6, 12]

APPENDIX A: PROOF OF RESULTS IN SECTION 3

For a vNM utility function $u : X \to [0,1]$ and consumption c_t, I use $u_{c_t} = [u(c_{t\omega})]_{\omega \in \Omega} \in \mathbb{R}^{\Omega}$ to denote the "utility act" corresponding to c_t. I use ζ, η to denote generic elements in $[0,1]^{\Omega}$. This section is organized as follows. Section A.1 states the proof of Proposition 3.1. Section A.2 states and proves an intermediate result used in subsequent sections. Section A.3 states the proof of Theorem 3.2, the main result of this paper. Section A.4 states the proof of Theorem 3.1. Section A.5 states the proof of Theorem 3.3.

A.1 *Proof of Proposition 3.1*

Fix parameter $\kappa = (\mu_1, u, \delta, \theta) \in K^*$, period-1 action $a^* = (c_1^*, C_2^*)$. To facilitate the proof, I first introduce Lemma A.1 and A.2.

LEMMA A.1. *For $c_2^* \in C_2^*$, the consumption stream $(c_1^*, c_2^*) \in \gamma_\kappa(a^*)$ if and only if for some $\mu^* \in \Delta\Omega$, (μ^*, c_2^*) is a solution to the optimization problem*

$$\max_{(\mu, c_2) \in \Delta\Omega \times C_2} \theta U_\mu(c_1^*, c_2) - D_{KL}(\mu || \mu_1). \tag{16}$$

PROOF. The joint optimization problem (16) can be solved sequentially. Specifically, the belief-consumption pair (μ^*, c_2^*) solves (16) if and only if both (17) and (18) are true:

$$\mu^* \in \arg\max_{\mu \in \Delta\Omega} \left\{ \max_{c_2 \in C_2} \theta U_\mu(c_1^*, c_2) - D_{KL}(\mu || \mu_1) \right\}, \tag{17}$$

$$c_2^* \in \arg\max_{c_2 \in C_2} \mathbb{E}_{\mu^*}[u(c_2)]. \tag{18}$$

Now, I prove the "if" side of Lemma A.1. Suppose $(c_1^*, c_2^*) \in \gamma_\kappa(a^*)$, then there exists a belief $\mu^* \in \Delta\Omega$, such that the intermediate player chooses μ^* and Player 2 chooses c_2^* in an SPNE of the CD game with menu $\{a^*\}$ and parameter κ. By the definition of SPNE of CD games (Definition 2.1), μ^* must solve (17) and c_2^* solves (18), which implies that the belief-consumption pair (μ^*, c_2^*) solves (16), the desired result.

For the "only if" side of Lemma A.1, suppose there exists $\mu^* \in \Delta\Omega$ such that (μ^*, c_2^*) solves (16). It follows that μ^* solves (17). By backward induction, according to Definition 2.1, μ^* is chosen by the intermediate player in some SPNE of the CD game with menu $\{a^*\}$ and parameter κ. According to (18), c_2^* maximizes the subjective expected utility with equilibrium belief μ^*. As a result, c_2^* is the period-2 action for some SPNE of the CD game with menu $\{a^*\}$ and parameter κ. That is, $(c_1^*, c_2^*) \in \gamma_\kappa(a^*)$. $\square$

LEMMA A.2. *For $c_2^* \in C_2^*$,*

$$c_2^* \in \arg\max_{c_2 \in C_2^*} \max_{\mu \in \Delta\Omega} \theta U_\mu(c_1^*, c_2) - D_{KL}(\mu || \mu_1) \tag{19}$$

if and only if

$$c_2^* \in \arg\max_{c_2 \in C_2^*} \mathbb{E}_{\mu_1}[\phi_\theta(u(c_1^*) + \delta u(c_2))]. \tag{20}$$

PROOF. Recall that $U_\mu(c_1^*, c_2) = \mathbb{E}_\mu[u(c_1^*) + \delta u(c_2)]$. Substitute this into the expression above, and Lemma A.2 follows directly from the variational formula in Dupuis and Ellis (1997). $\square$

Proposition 3.1 directly follows from Lemma A.1 and A.2. For $c_2^* \in C_2^*$, Lemma A.1 states that $(c_1^*, c_2^*) \in \gamma_\kappa(a^*)$ if and only if for some $\mu^* \in \Delta\Omega$, (μ^*, c_2^*) solves (16). This, according to Lemma A.2, is equivalent to (20), the desired result.

A.2 *An Intermediate Result*

In this section, I present Theorem A.1, an intermediate result used in the proofs of both Theorem 3.1 and Theorem 3.2. Recall that $\mathcal{A}_1$ (resp. $\mathcal{A}_2$) represents the collection of 1-determined (resp. 2-determined) menus, where the consumption choice is determined solely by the period-1 (resp. period-2) self. Theorem A.1 establishes that choices over the domains $\mathcal{A}_1$ and $\mathcal{A}_2$ are made "as if" they maximize certain utility functions corresponding to each consumption stream.

THEOREM A.1. *An induced choice correspondence γ satisfies Axioms 1 and 2 in the domain $\mathcal{A}_1$, if and only if there exists $\mu_1 \in Int(\Delta\Omega)$, $u \in \mathcal{U}^*$, $\delta \in (0,1)$, and a family of continuous and strictly monotone functional $\{I_{c_1}\}_{c_1 \in \mathcal{C}_1}$ over $[0,1]^\Omega$ such that for all $A \in \mathcal{A}_1$,*

$$\gamma(A) = \arg\max_{c \in A} \mathbb{E}_{\mu_1}[u(c_1)] + \delta \mathbb{E}_{\mu_1}[u(c_2)]; \tag{21}$$

Moreover, if γ satisfies Axioms 1, 2, 4, 5 in the domain $\mathcal{A}_2$, then for all $c_1 \in \mathcal{C}_1$ and $a = (c_1, C_2)$,

$$\gamma_2(a) = \arg\max_{c_2 \in C_2} I_{c_1}[u(c_2)]. \tag{22}$$

PROOF. I start with proving the first half of Theorem A.1. The "if" direction is obvious. To prove the "only if" direction, I Fix an induced choice correspondence $\gamma : \mathcal{A} \rightsquigarrow \mathcal{C}$ that satisfies Axioms 1 and 2. Consider the restriction of γ to the domain $\mathcal{A}_1$. By Axiom 1 (i) (WARP), the restriction of γ over $\mathcal{A}_1$ induces weak order $\succsim^*$ over $\mathcal{C}$ such that for every $A \in \mathcal{A}_1$ and $c \in A$,

$$c \in \gamma(A) \iff c \succsim^* c' \text{ for every } c' \in A.$$

Lemma A.3 below states the basic properties of the preference relation $\succsim^*$. For $x \in X$, $t \in \mathcal{T}$, $\omega \in \Omega$ and $c, c' \in \mathcal{C}$, we say $c' = x^c_{t\omega}$ if

$$c'_{\tau\psi} = \begin{cases} x & \text{if } \tau = t, \psi = \omega; \\ c_{\tau\psi} & \text{otherwise.} \end{cases}$$

LEMMA A.3. *For $c, c', c'' \in \mathcal{C}$, the binary relation $\succsim^*$ satisfies the following properties:*

1. *there exists c, c' such that $c \succ^* c'$;*
2. *if $c \succ^* c'$ and $\alpha \in (0,1)$, then $\alpha c + (1-\alpha)c'' \succ^* \alpha c' + (1-\alpha)c''$;*
3. *$\succsim^*$ is continuous in the sense that the upper and lower contour sets are closed;*
4. *for $x, y \in X$, $s, t \in \mathcal{T}$ and $\omega, \psi \in \Omega$,*

$$x^c_{t\omega} \succsim^* y^c_{t\omega} \iff x^{c'}_{\tau\psi} \succsim^* y^{c'}_{\tau\psi} \tag{23}$$

PROOF. The lemma is a direct consequence of applying Axiom 1 to the domain $\mathcal{A}_1$.

- For Lemma A.3 part 1, assume for the sake of contradiction that the statement is false. If this were the case, then all consumption streams would be indifferent to each other. Consequently, for every $A \in \mathcal{A}_1$, it holds that $\gamma(A) = A$. This is contradictory to Axiom 1 (ii).
- For Lemma A.3 part 2, Axiom 1 (iii) implies that if $\gamma(\{c, c'\}) = c$, then $\gamma(\{\alpha c + (1-\alpha)c'', \alpha c' + (1-\alpha)c''\}) = \alpha c + (1-\alpha)c''$, which implies part 2.

- For Lemma A.3 part 3, to see that the upper contour sets are closed, fix $c \in \mathcal{C}$ and consider sequence $\{c^n\}_{n=1}^{\infty} \in \mathcal{C}^{\infty}$ such that $c^n \succsim^* c$ for every n and $c^n \to c'$ for some $c' \in \mathcal{C}$. Then $c^n \in \gamma(\{c, c^n\})$ for all n. By the definition of the Hausdorff metric, it holds that $\{c, c^n\} \to \{c, c'\}$. Therefore, Axiom 1 (iv) implies $\gamma(\{c, c^n\}) \to \gamma(\{c, c'\})$, which implies $c' \in \gamma(\{c, c'\})$, that is, $c' \succsim^* c$. As a result, the upper contour sets are closed. The proof that the lower contour sets are closed is similar.

- For Lemma A.3 part 4, take $L = \{x, y\}$ and (23) follows directly from Axiom 1 (v).

$\square$

By Lemma A.3, the binary relation $\succsim^*$ satisfies the Anscombe-Aumann axioms over the domain $\mathcal{C}$, therefore, there exists a non-constant and linear utility function u over X and a probability measure $\hat{\mu} = [\hat{\mu}(t, \omega)]_{t \in \mathcal{T}, \omega \in \Omega}$ over $\mathcal{T} \times \Omega$ with full support such that the preference $\succsim^*$ is represented by $\hat{U} : \mathcal{C} \to \mathbb{R}$, where

$$\hat{U}(c) = \sum_{\omega \in \Omega} \hat{\mu}(1, \omega) u(c_{1\omega}) + \sum_{\omega \in \Omega} \hat{\mu}(2, \omega) u(c_{2\omega}). \tag{24}$$

LEMMA A.4. *There exists $\mu_1 = (\mu_{1\omega})_{\omega \in \Omega} \in Int(\Delta\Omega)$ such that $\succsim^*$ is represented by*

$$U_{\mu_1}(c) = \sum_{\omega \in \Omega} \mu_{1\omega} u(c_{1\omega}) + \delta \sum_{\omega \in \Omega} \mu_{1\omega} u(c_{2\omega}),$$

where $u(X) = [0, 1]$.

PROOF. By Axiom 2 (Discounting), $(c_1, c_1) \succ^* (c_1', c_1')$ implies $(c_1, c_1') \succ^* (c_1', c_1)$. Notice that according to (24), for $c_1, c_1' \in \mathcal{C}_1$, $(c_1, c_1) \succ^* (c_1', c_1')$ if and only if

$$\sum_{\omega \in \Omega} [\hat{\mu}(1, \omega) + \hat{\mu}(2, \omega)] u(c_{1\omega}) > \sum_{\omega} [\hat{\mu}(1, \omega) + \hat{\mu}(2, \omega)] u(c_{1\omega}'). \tag{25}$$

This implies

$$\sum_{\omega \in \Omega} \hat{\mu}(1, \omega) u(c_{1\omega}) + \sum_{\omega \in \Omega} \hat{\mu}(2, \omega) u(c_{1\omega}') > \sum_{\omega \in \Omega} \hat{\mu}(1, \omega) u(c_{1\omega}') + \sum_{\omega \in \Omega} \hat{\mu}(2, \omega) u(c_{1\omega}),$$

which can be rewritten as

$$\sum_{\omega}(\hat{\mu}[1,\omega)-\hat{\mu}(2,\omega)]u(c_{1\omega})>\sum_{\omega}[\hat{\mu}(1,\omega)-\hat{\mu}(2,\omega)]u(c'_{1\omega}). \tag{26}$$

Notice that (25) and (26) apply to arbitrarily chosen consumptions $c_1, c'_1 \in \mathcal{C}_1$. Apply the uniqueness of expected utility representations to (25) and (26), it follows that there exists a $\delta_0 > 0$ such that $\hat{\mu}(1,\omega)+\hat{\mu}(2,\omega)=\delta_0(\hat{\mu}(1,\omega)-\hat{\mu}(2,\omega))$ for all $\omega \in \Omega$. Moreover, since $\hat{\mu}(2,\omega) > 0$ for all ω, it must be that $\delta_0 > 1$. Denote $\delta = (\delta_0-1)/(\delta_0+1) \in (0,1)$, then it follows that $\hat{\mu}(2,\omega) = \delta\hat{\mu}(1,\omega)$ for all $\omega \in \Omega$. Define $\mu_1 = (\mu_{1\omega})_{\omega\in\Omega} \in Int(\Delta\Omega)$ such that $\mu_{1\omega} = \hat{\mu}(1,\omega)/\sum_{\omega'\in\Omega}\hat{\mu}(1,\omega')$ for all $\omega \in \Omega$, then the representation $\hat{U}$ of $\succsim^*$ in (24) is ordinally equivalent to the utility function U_{μ_1} defined as follows:

$$U_{\mu_1}(c) = \sum_{\omega\in\Omega}\mu_{1\omega}u(c_{1\omega}) + \delta\sum_{\omega\in\Omega}\mu_{1\omega}u(c_{2\omega}), \tag{27}$$

where $\delta \in (0,1)$ and $\mu_1 \in Int(\Delta\Omega)$. Without loss of generality, we may assume $u(X) = [0,1]$. $\square$

Lemma A.4 above establishes the first half of Theorem A.1. For the second half of the theorem, it remains to show that Axioms 1, 2, 4, 5 implies (22). That is, fix any $\hat{c}_1 \in \mathcal{C}_1$, there exists a continuous and strictly monotone functional $I_{\hat{c}_1} : [0,1]^\Omega \to \mathbb{R}$ such that for finite $C_2 \subset \mathcal{C}_2$,

$$\gamma_2[(\hat{c}_1, C_2)] = \arg\max_{c_2\in C_2} I_{\hat{c}_1}[u(c_2)].$$

First, Recall that γ_2 reflects the period-2 consumption corresponding to the induced choice correspondence γ. Therefore, $\gamma_2[(\hat{c}_1, C_2)]$ captures the period-2 consumption resulting from the singleton menu $\{(\hat{c}_1, C_2)\}$, which is an element of $\mathcal{A}_2$. Apply Axiom 1 (i) (WARP) to the restricted domain of $\mathcal{A}_2$, then there exists a weak order $\succsim^*_{\hat{c}_1}$ over $\mathcal{C}_2$ such that for every finite $C_2 \subset \mathcal{C}_2$,

$$c_2 \in \gamma_2[(\hat{c}_1, C_2)] \iff c_2 \succsim^*_{\hat{c}_1} c'_2 \text{ for every } c'_2 \in C_2.$$

Moreover, by Axiom 1 (iv), $\succsim^*_{\hat{c}_1}$ is continuous in the sense that the upper and lower contour sets are closed.[36] The next lemma shows that the DM's taste remains stable over time; specifically, if payoff x is preferred to y according to the period-1 utility index u, the period-2 self will also choose x over y.

LEMMA A.5. *Let $\hat{c}_1 \in \mathcal{C}_1$. For the utility index u in Lemma A.4, for $x, y \in X$, $\hat{c}_2 \in \mathcal{C}_2$, $\omega \in \Omega$ and $E = \{\omega\}$,*

$$u(x) \geq u(y) \iff xE\hat{c}_2 \succsim^*_{\hat{c}_1} yE\hat{c}_2.$$

PROOF. Here, the utility index $u : X \to [0,1]$ represents the DM's taste over X *ex-ante*. Lemma A.5 imposes the minimal dynamic consistency requirement that the DM's taste does not change over time. The proof of Lemma A.5 will use Axiom 4, which imposes a weakened dynamic consistency condition on the DM's choices. To start the proof, consider the payoff $z \in X$ such that $z = \hat{c}_{1\omega}$. I construct $\tilde{c} = (z, xEy)$, $\tilde{c}' = (z, y)$ and $\tilde{a} = (z, \{xEy, y\})$. Also, let $\hat{c} = (\hat{c}_1, \hat{c}_2)$. In this case,

$$\begin{aligned} u(x) \geq u(y) &\iff \tilde{c} \in \gamma(\{\tilde{c}, \tilde{c}'\}) \\ &\iff \tilde{c} \in \gamma(\tilde{a}) && (*) \\ &\iff \tilde{c}E\hat{c} \in \gamma(\tilde{a}E\hat{c}) && (**) \\ &\iff (\hat{c}_1, xE\hat{c}_2) \in \gamma[(\hat{c}_1, \{xE\hat{c}_2, yE\hat{c}_2\})] && (***) \\ &\iff xE\hat{c}_2 \succsim^*_{\hat{c}_1} yE\hat{c}_2. \end{aligned}$$

In the sequence of statements above, $(*)$ follows from Axiom 4 (Cognitive Consistency). To see the reason why, first notice that event E is desirable for $\tilde{a}$ because x is (weakly) preferred over y *ex-ante*. Then, notice that $\tilde{c}$ is preferred over $\tilde{c}'$ *ex-ante* as $u(x) \geq u(y)$. Moreover, $\tilde{c}$ is still preferred over $\tilde{c}'$ conditional on the desirable event E, as the payoff stream in E of $\tilde{c}$ is (z, x), which is *ex-ante* preferred over (z, y). As a result, Axiom 4 implies that $\tilde{c} \in \gamma(\tilde{a})$, as is stated in $(*)$. Then, $(**)$ directly follows from Axiom 5 (STP). Finally, $(***)$ rewrites the consumption stream $\tilde{c}E\hat{c}$ and action $\tilde{a}E\hat{c}$ in $(**)$ according to how these objects are constructed. □

[36] Consider sequence $\{c_2^n\}_{n=1}^{\infty} \in \mathcal{C}_2^{\infty}$ such that $c_2^n \to c_2'$ and $c_2^n \succsim^*_{\hat{c}_1} c_2$ for every n. Then $(\hat{c}_1, c_2^n) \in \gamma[(\hat{c}_1, \{c_2, c_2^n\})]$ for every n. By the definition of Hausdorff metric, it holds that $\{(\hat{c}_1, \{c_2, c_2^n\})\} \to \{(\hat{c}_1, \{c_2, c_2'\})\}$. Axiom 1 (iv) implies $(\hat{c}_1, c_2') \in c[(\hat{c}_1, \{c_2, c_2'\})]$, that is, $c_2' \succsim_{\hat{c}_1} c_2$. As a result, the upper contour sets are closed. The proof that the lower contour sets are closed is analogous.

As a result of Lemma A.5 and Axiom 5 (STP), for $c_2, c_2' \in \mathcal{C}_2$, $c_2 \succsim^*_{\hat{c}_1} c_2'$ if $u(c_{2\omega}) \geq u(c_{2\omega}')$ for all $\omega \in \Omega$ and $u(c_{2\psi}) > u(c_{2\psi}')$ for some $\psi \in \Omega$. Moreover, recall from the analysis above that $\succsim^*_{\hat{c}_1}$ is a continuous weak order over $\mathcal{C}_2$. Therefore, there exists a continuous and strictly monotone functional $I_{\hat{c}_1} : [0,1]^\Omega \to \mathbb{R}$ such that $\succsim^*_{\hat{c}_1}$ is represented by $I_{\hat{c}_1}[u(c_2)]$ over $\mathcal{C}_2$. □

A.3 *Proof of Theorem 3.2*

Before starting the proof, note that the uniqueness result in Theorem 3.2 follows directly from Strzalecki (2011). In the remainder of this section, I present the proof of the "only if" direction of Theorem 3.2; the "if" direction follows from standard arguments. For the family of functionals $[I_{c_1}(\zeta)]_{c_1 \in \mathcal{C}_1}$ constructed in Theorem A.1, I want to show that there exists $\theta \geq 0$, such that for every $c_1 \in \mathcal{C}_1$, the functional $I_{c_1}(\zeta)$ is ordinally equivalent to $\mathbb{E}_{\mu_1}[\phi_\theta(u(c_1) + \delta\zeta)]$. Let $\trianglerighteq_{c_1}$ denote the binary relation over $[0,1]^\Omega$ represented by $I_{c_1}(\cdot)$. For $\zeta, \eta \in [0,1]^\Omega$, we say $\zeta \equiv_{c_1} \eta$ if $\zeta \trianglerighteq_{c_1} \eta$ and $\eta \trianglerighteq_{c_1} \zeta$.

LEMMA A.6. *For every $E \in \mathcal{E}$, $c_1, c_1' \in \mathcal{C}_1$, and $\zeta, \eta, \zeta' \in [0,1]^\Omega$,*

(i) $\zeta \trianglerighteq_{c_1} \eta E \zeta \implies \zeta E \zeta' \trianglerighteq_{c_1} \eta E \zeta'$*; and*

(ii) $\zeta \trianglerighteq_{c_1} \eta E \zeta \implies \zeta \trianglerighteq_{c_1 E c_1'} \eta E \zeta$.

Moreover, for every $h \in [-1,1]$ such that $\zeta + h 1_\Omega, \eta + h 1_\Omega \in [0,1]^\Omega$,

$$\zeta \trianglerighteq_{c_1} \eta \iff \zeta + h 1_\Omega \trianglerighteq_{c_1} \eta + h 1_\Omega. \tag{28}$$

PROOF. Fix $E \in \mathcal{E}$, $c_1, c_1' \in \mathcal{C}_1$, and $\zeta, \eta, \zeta' \in [0,1]^\Omega$. Let $\hat{c}_2, \hat{\hat{c}}_2, \hat{c}_2' \in \mathcal{C}_2$ be such that $u(\hat{c}_2) = \zeta$, $u(\hat{\hat{c}}_2) = \eta$ and $u(\hat{c}_2') = \zeta'$. To prove Lemma A.6 (i), it suffices to show that

$$\hat{c}_2 \succsim^*_{c_1} \hat{\hat{c}}_2 E \hat{c}_2 \implies \hat{c}_2 E \hat{c}_2' \succsim^*_{c_1} \hat{\hat{c}}_2 E \hat{c}_2'.$$

This is equivalent to show

$$(c_1, \hat{c}_2) \in \gamma[(c_1, \{\hat{c}_2, \hat{\hat{c}}_2 E \hat{c}_2\})] \implies (c_1, \hat{c}_2 E \hat{c}_2') \in \gamma[(c_1, \{\hat{c}_2 E \hat{c}_2', \hat{\hat{c}}_2 E \hat{c}_2'\})].$$

Define $\hat{A} = \{(c_1, \{\hat{c}_2, \hat{\hat{c}}_2 E \hat{c}_2\}))\}$ and $\hat{c} = (c_1, \hat{c}_2')$, then the expression above states that $(c_1, \hat{c}_2) \in \gamma(\hat{A}) \implies (c_1, \hat{c}_2 E \hat{c}_2') \in \gamma(\hat{A} E \hat{c})$, a direct consequence of Axiom 5 (STP).

Now I prove Lemma A.6 (ii). It suffices to show that

$$\hat{c}_2 \succsim^*_{c_1} \hat{\hat{c}}_2 E \hat{c}_2 \implies \hat{c}_2 \succsim^*_{c_1 E c_1'} \hat{\hat{c}}_2 E \hat{c}_2.$$

This is equivalent to show

$$(c_1, \hat{c}_2) \in \gamma[(c_1, \{\hat{c}_2, \hat{\hat{c}}_2 E \hat{c}_2\})] \implies (c_1 E c_1', \hat{c}_2) \in \gamma[(c_1 E c_1', \{\hat{c}_2, \hat{\hat{c}}_2 E \hat{c}_2\})].$$

Define $\hat{\hat{A}} = \{(c_1, \{\hat{c}_2, \hat{\hat{c}}_2 E \hat{c}_2\})\}$ and $\hat{\hat{c}} = (c_1', \hat{c}_2)$, then the expression above states that $(c_1, \hat{c}_2) \in \gamma(\hat{\hat{A}}) \implies (c_1 E c_1', \hat{c}_2) \in \gamma(\hat{\hat{A}} E \hat{\hat{c}})$, a direct consequence of Axiom 5 (STP).

Finally, fix $h \in [-1, 1]$ such that $\zeta + h 1_\Omega$ and $\eta + h 1_\Omega \in [0,1]^\Omega$. Without loss of generality assume $h > 0$. I want to prove that (28) is true. First, I take $\underline{h} = \min \cup_{\omega \in \Omega} \{\eta_\omega, \zeta_\omega\}$ and $\bar{h} = \max \cup_{\omega \in \Omega} \{\eta_\omega, \zeta_\omega\}$. Take $\Delta = \bar{h} - \underline{h}$. I construct the following two elements of $[0,1]^\Omega$:

$$\begin{cases} \tilde{\zeta} = \left(1 + \frac{h}{\Delta}\right) \zeta - \frac{h}{\Delta} \underline{h}, \\ \tilde{\eta} = \left(1 + \frac{h}{\Delta}\right) \eta - \frac{h}{\Delta} \underline{h}; \end{cases}$$

then take $\hat{\alpha} = \frac{\Delta}{\Delta + h}$, it holds that

$$\begin{cases} \zeta = \hat{\alpha} \tilde{\zeta} + (1 - \hat{\alpha}) \underline{h}, & \eta = \hat{\alpha} \tilde{\eta} + (1 - \hat{\alpha}) \underline{h}; \\ \zeta + h = \hat{\alpha} \tilde{\zeta} + (1 - \hat{\alpha})(\bar{h} + h), & \eta + h = \hat{\alpha} \tilde{\eta} + (1 - \hat{\alpha})(\bar{h} + h). \end{cases}$$

Take period-2 consumption $d_2, d_2' \in \mathcal{C}_2$ such that $u(d_2) = \tilde{\zeta}$ and $u(d_2') = \tilde{\eta}$. Take payoffs $\hat{x}, \hat{y} \in X$ such that $u(\hat{x}) = \underline{h}$ and $u(\hat{y}) = \bar{h} + h$. Then Axiom 6 (Weak C-Independence) implies

$$\hat{\alpha} d_2 + (1 - \hat{\alpha}) \hat{x} \succsim^*_{c_1} \hat{\alpha} d_2' + (1 - \hat{\alpha}) \hat{x} \implies \hat{\alpha} d_2 + (1 - \hat{\alpha}) \hat{y} \succsim^*_{c_1} \hat{\alpha} d_2' + (1 - \hat{\alpha}) \hat{y},$$

which, according to how the mathematical objects above are constructed, implies the "only if" side of (28). The proof for the "if" side is analagous. □

LEMMA A.7. *For every $c_1 \in \mathcal{C}_1$, there exists continuous and strictly increasing functions $\{\phi_{c_1,\omega}\}_{\omega \in \Omega} : [0,1] \to \mathbb{R}$ such that $\trianglerighteq_{c_1}$ is represented by I'_{c_1}, which is defined as*

$$I'_{c_1}(\zeta) = \sum_\omega \phi_{c_1,\omega}(\zeta_\omega).$$

PROOF. Fix $c_1 \in \mathcal{C}_1$. Lemma A.6 (i) implies

$$\zeta \trianglerighteq_{c_1} \eta E\zeta \implies \zeta E\zeta' \trianglerighteq_{c_1} \eta E\zeta'.$$

Moreover, $\trianglerighteq_{c_1}$ is continuous as it has a continous representation. Therefore, by Theorem 3 of Debreu (1959), there exists continuous functions $\{\phi_{c_1,\omega}\}_{\omega\in\Omega} : [0,1] \to \mathbb{R}$ such that $\trianglerighteq_{c_1}$ is represented by

$$I'_{c_1}(\zeta) = \sum_{\omega} \phi_{c_1,\omega}(\zeta_\omega).$$

□

LEMMA A.8. *For every $x \in X$, there exists a strictly increasing, continuous and convex function ϕ_x such that $\trianglerighteq_x$ is ordinally equivalent to $\mathbb{E}_{\mu_1}[\phi_x(\zeta)]$.*

PROOF. Fix an $x \in X$ and an arbitrary $\hat{\omega} \in \Omega$. Define $\phi_{x,-\hat{\omega}}(\cdot) = \sum_{\omega\neq\hat{\omega}} \phi_{x,\omega}(\cdot)$, and define $J_{x,\hat{\omega}} : [0,1]^2 \to \mathbb{R}$ as

$$J_{x,\hat{\omega}}(b,b') = \phi_{x,\hat{\omega}}(b) + \phi_{x,-\hat{\omega}}(b'). \tag{29}$$

I first claim that for all $b, b' \in [0,1]$, it holds that

$$J_{x,\hat{\omega}}(b,b') \geq J_{x,\hat{\omega}}[\mu_{1\hat{\omega}}b + (1-\mu_{1\hat{\omega}})b', \mu_{1\hat{\omega}}b + (1-\mu_{1\hat{\omega}})b']. \tag{30}$$

Fix $b, b' \in [0,1]$. Consider a period-2 consumption $\hat{c}_2 \in \mathcal{C}_2$ such that $u(\hat{c}_{2\hat{\omega}}) = b$ and $u(\hat{c}_{2\omega}) = b'$ for all $\omega \neq \hat{\omega}$. For each $\varepsilon < 0$, define the payoff $y^\epsilon \in X$ such that $u(y^\epsilon) = \mu_{1\hat{\omega}}b + (1-\mu_{1\hat{\omega}})b' - \epsilon$. Let a^ϵ define the period-1 action $(x, \{\hat{c}_2, y^\epsilon\})$. Fix an arbitrary $\epsilon > 0$. By construction, the *ex-ante* expected utility of $\hat{c}_2$ is strictly greater than the utility of y^ϵ. As a result,

$$\gamma[F(a^\epsilon)] = (x, \hat{c}_2). \tag{31}$$

Next, I show that (30) holds by discussing the three cases below.

- Case 1: $b > b'$; in this case, take $E = \{\hat{\omega}\}$, and E is desirable for a^ϵ according to Definition 3.4. Moreover, $u(\hat{c}_{2\hat{\omega}}) = b > \mu_{1\hat{\omega}}b + (1-\mu_{1\hat{\omega}})b' > u(y^\epsilon)$. As a result,

$$\gamma\left[F_E(a^\epsilon)\right] = (x, \hat{c}_{2\hat{\omega}}). \tag{32}$$

By (31), (32) and Axiom 4 (Cognitive Consistency), $\gamma(a^\epsilon) = (x, \hat{c}_2)$. As a result, $I_x[u(\hat{c}_2)] \geq I_x[u(y^\epsilon)] = u(y^\epsilon)$, [37] where $u(y^\epsilon) = \mu_{1\hat{\omega}} b + (1 - \mu_{1\hat{\omega}})b' - \epsilon$ by definition. Take $\epsilon \to 0$, and $I_x[u(\hat{c}_2)] \geq I_x[(\mu_{1\hat{\omega}} b + (1 - \mu_{1\hat{\omega}})b') \cdot 1_\Omega]$. By the construction of $\hat{c}_2$, the left-hand side of the inequality equals to $J_{x,\hat{\omega}}(b, b')$, which implies (30).

- Case 2: $b < b'$; in this case, take $G = \Omega - \{\hat{\omega}\}$, and G is desirable for a^ϵ according to Definition 3.4. Moreover, $u(\hat{c}_{2G}) = b' > \mu_{1\hat{\omega}} b + (1 - \mu_{1\hat{\omega}})b' > u(y^\epsilon)$. As a result,

$$\gamma\left[F_G(a^\epsilon)\right] = (x, \hat{c}_{2G}). \tag{33}$$

  By (31) and (33) and Axiom 4 (Cognitive Consistency), $\gamma(a^\epsilon) = (x, \hat{c}_2)$. The remainder of the proof showing how the equality implies (30) is identical to the reasoning in Case 1.

- Case 3: $b = b'$; in the analysis above, we show that $J_{x,\hat{\omega}}$ satisfies (30) if $b > b'$ or $b < b'$. The continuity of $J_{x,\hat{\omega}}$ implies that (30) holds if $b = b'$.

According to inequality (30), by Werner (2005), the $\phi_{x,\hat{\omega}}$ and $\phi_{x,-\hat{\omega}}$ defined in (29) are convex and there exists a constant $h_{\hat{\omega}} \in \mathbb{R}$ such that

$$\phi_{x,\hat{\omega}} = \frac{\mu_{1\hat{\omega}}}{1 - \mu_{1\hat{\omega}}} \phi_{x,-\hat{\omega}} + h_{\hat{\omega}}.$$

Therefore, $(1 - \mu_{1\hat{\omega}})\phi_{x,\hat{\omega}} = \mu_{1\hat{\omega}}\phi_{x,-\hat{\omega}} + (1 - \mu_{1\hat{\omega}})h_{\hat{\omega}}$. Recall $\phi_{x,-\hat{\omega}} = \sum_{\omega \neq \hat{\omega}} \phi_{x,\omega}$. Add $\mu_{1\hat{\omega}}$ on both sides of the equation, we have $\phi_{x,\hat{\omega}} = \mu_{1\hat{\omega}} \sum_{\omega \in \Omega} \phi_{x,\omega} + (1 - \mu_{1\hat{\omega}})h_{\hat{\omega}}$. Notice that the term $\sum_{\omega \in \Omega} \phi_{x,\omega}$ is independent of the choice of $\hat{\omega}$. Denote $\sum_\omega \phi_{x,\omega}$ as ϕ_x. Since $\hat{\omega}$ is arbitrarily chosen, for every $\omega' \in \Omega$, there exists $h_{\omega'} \in \mathbb{R}$ such that $\phi_{x,\omega'} = \mu_{1\omega'}\phi_x + (1 - \mu_{1\omega'})c_{\omega'}$. As a result, $I'_x(\zeta) = \sum_{\omega \in \Omega} \mu_{1\omega}\phi_x(\zeta_\omega) + \sum_{\omega \in \Omega}(1 - \mu_{1\omega})h_\omega$, which is ordinally equivalent to $\sum_{\omega \in \Omega} \mu_{1\omega}\phi_x(\zeta_\omega)$, *i.e.* $\mathbb{E}_{\mu_1}[\phi_x(\zeta)]$. The function ϕ_x must be strictly increasing and continuous as I_x is strictly monotonic and continuous over $[0,1]^\Omega$. □

LEMMA A.9. *For every $x \in X$, there exists a $\theta(x) \geq 0$ such that $\trianglerighteq_x$ is represented by* $\mathbb{E}_{\mu_1}[\phi_{\theta(x)}(\zeta)]$.

PROOF. Fix $x \in X$. By Lemma A.8 above, $\trianglerighteq_x$ is represented by $\mathbb{E}_{\mu_1}[\phi_x(\zeta)]$, where ϕ_x is strictly increasing. Therefore, $\trianglerighteq_x$ is a continuous and strictly monotone weak order, which satisfies Axioms A1, A3, A4, and A6 in Strzalecki (2011). Since the state space Ω is finite, Axiom A8

[37] The last equation holds because at the start of Section A.3, I assumed without loss of generality that $I_{c_1}(h1_\Omega) = h$ for all $c_1 \in \mathcal{C}_1$ and $h \in [0,1]$.

in Strzalecki (2011) is also satisfied. Moreover, since ϕ_x is convex, the order $\trianglerighteq_x$ satisfies the following *convexity* condition:

$$\zeta \equiv_x \eta \implies \zeta \trianglerighteq_x \alpha\zeta + (1-\alpha)\eta. \tag{34}$$

Also, recall that Lemma A.6 states that the following two statements are true:

$$\zeta \trianglerighteq_x \eta E \zeta \implies \zeta E \zeta' \trianglerighteq_x \eta E \zeta', \tag{35}$$

and

$$\zeta \trianglerighteq_x \eta \iff \zeta + h 1_\Omega \trianglerighteq_x \eta + h 1_\Omega. \tag{36}$$

Statements (34), (35), (36) correspond to Axiom A5, P2 and A2 in Strzalecki (2011) respectively.[38] Lemma A.9 follows from a direct application of Theorem 1 in Strzalecki (2011). □

DEFINITION A.1. *For $c_1 \in \mathcal{C}_1$, $\zeta, \eta \in [0,1]^\Omega$ and $E \in \mathcal{E}$, we say $\zeta \trianglerighteq_{c_1}^E \eta$ if for all ζ',*

$$\zeta E \zeta' \trianglerighteq_{c_1} \eta E \zeta'.$$

By Lemma A.6 (i), the order $\trianglerighteq_{c_1}^E$ is well-defined.

LEMMA A.10. *Let $\omega^* \in \Omega$ and $c_1, c_1' \in \mathcal{C}_1$ such that $c_{1\omega} = c'_{1\omega}$ for every $\omega \neq \omega^*$. If there exists $f_{c_1} : \Omega \to \mathbb{R}_{++}$ and $\theta(c_1) \geq 0$ such that $\trianglerighteq_{c_1}$ is represented by $\mathbb{E}_{\mu_1}[f_{c_1}(\omega) \cdot \phi_{\theta(c_1)}(\zeta_\omega)]$, then there exists $f_{c_1'} : \Omega \to \mathbb{R}_{++}$ and $\theta(c_1') \geq 0$, such that $\trianglerighteq_{c_1}$ is represented by*

$$\mathbb{E}_{\mu_1}[f_{c_1'}(\omega) \cdot \phi_{\theta(c_1')}(\zeta_\omega)].$$

Moreover, $\theta(c_1) = \theta(c_1')$.

PROOF. Fix $F = \Omega - \{\omega^*\}$. By Lemma A.6 (ii), it holds that $\trianglerighteq_{c_1}^F = \trianglerighteq_{c_1'}^F$. By Lemma A.7, there exists continuous and strictly increasing functions $\{\phi_{c_1',\omega}\}_{\omega\in\Omega}$ such that $\trianglerighteq_{c_1'}$ is represented by $\sum_{\omega\in\Omega} \phi_{c_1',\omega}(\zeta_\omega)$. As a result, $\trianglerighteq_{c_1'}^F$ is represented by $\sum_{\omega\in F} \phi_{c_1',\omega}(\zeta_\omega)$. On the other hand, $\trianglerighteq_{c_1}^F$

[38] Expression (34) asserts that the *lower contour set* corresponding to $\trianglerighteq_x$ is convex. In contrast, Axiom A5 in Strzalecki (2011) (which is the uncertainty aversion axiom in Gilboa and Schmeidler (1989)) asserts that the *upper contour set* corresponding to the binary relation is convex. Nonetheless, the main representation theorem in Strzalecki (2011) remains applicable, with the distinction that the resulting representation in my case is convex, whereas it is concave in his case.

is represented by $\sum_{\omega\in F}\mu_{1\omega}\cdot f_{c_1}(\omega)\cdot\phi_{\theta(c_1)}(\zeta_\omega)$. Since $\trianglerighteq^F_{c_1}=\trianglerighteq^F_{c_1'}$, by the uniqueness property of additive utility representations, there exists $k>0$ and $b\in\mathbb{R}$ such that $\phi_{c_1',\omega}=k\mu_{1\omega}\cdot f_{c_1}(\omega)\cdot\phi_{\theta(c_1)}+b$ for every $\omega\in F$. Now denote $\phi_{c_1',\omega^*}=\Phi:[0,1]\to\mathbb{R}$, then $\trianglerighteq_{c_1'}$ is represented by

$$\hat{I}_{c_1'}(\zeta)\equiv\Phi(\zeta_{\omega^*})+\sum_{\omega\neq\omega^*}k\mu_{1\omega}\cdot f_{c_1}(\omega)\cdot\phi_{\theta(c_1)}(\zeta_\omega)+b. \tag{37}$$

The $\hat{I}_{c_1'}$ defined above is a functional $[0,1]^\Omega\to\mathbb{R}$. Now I want to prove that, for some $k'>0$ and $b'\in\mathbb{R}$,

$$\Phi=k'\phi_{\theta(c_1)}+b'. \tag{38}$$

Without loss of generality, assume $\Omega=\{1,2,...,n\}$ where $n\geq 3$, and $\omega^*=1$. To prove that (38) holds for some $k'>0$ and b, I first claim that for any $h,h'\in(0,1)$ such that $h+h'\in(0,1)$,

$$\frac{\Phi(h+h')-\Phi(h')}{\Phi(h)-\Phi(0)}=\exp\left[\theta(c_1)\cdot h'\right]. \tag{39}$$

To prove (39), fix $h,h'\in(0,1)$ such that $h+h'\in(0,1)$. Take $K>0$ sufficiently large such that

$$\Phi\left(\hat{h}+\frac{h}{K}\right)-\Phi(\hat{h})<\sum_{\omega\neq\omega^*}\mu_{1\omega}\cdot f_{c_1}(\omega)\cdot[\phi_{\theta(c_1)}(1-h')-\phi_{\theta(c_1)}(0)]. \tag{40}$$

for all $\hat{h},\hat{h}+\frac{h}{K}\in(0,1)$. Notice the right-hand side of (40) is strictly positive since $\mu_1\in Int(\Delta\Omega)$, $f(\cdot)>0$ and $h'\in(0,1)$. Such K exists because Φ is a continuous function over the compact set $[0,1]$, which implies the uniform continuity of Φ by the Heine–Cantor theorem. For all $k_1=0,1,...,K-1$, let $\Delta_{k_1}\in(0,1-h')$ be the unique number such that

$$\Phi\left(\frac{k_1}{K}h\right)+\sum_{\omega\neq\omega^*}\mu_{1\omega}\cdot f_{c_1}(\omega)\cdot\phi_{\theta(c_1)}(\Delta_{k_1})=\Phi\left(\frac{k_1+1}{K}h\right)+\sum_{\omega\neq\omega^*}\mu_{1\omega}\cdot f_{c_1}(\omega)\cdot\phi_{\theta(c_1)}(0). \tag{41}$$

First, $\Delta_{k_1}>0$ exists and is unique by the continuity and strict monotonicity of $\Phi(\cdot)$ and $\phi_{\theta(c_1)}(\cdot)$, as well as the unboundedness of $\phi_{\theta(c_1)}(\cdot)$. Second, I want to show $\Delta_{k_1}<1-h'$. Rearrange (41), and we get $\sum_{\omega\neq\omega^*}\mu_{1\omega}f_{c_1}(\omega)[\phi_{\theta(c_1)}(\Delta_{k_1})-\phi_{\theta(c_1)}(0)]=F\left(\frac{k_1+1}{K}h\right)-F\left(\frac{k_1}{K}h\right)$. Substitute the right-hand side of this equation by the right-hand side of (40), rearrange and we get

$$\phi_{\theta(c_1)}(\Delta_{k_1})<\phi_{\theta(c_1)}(1-h'),$$

which implies $\Delta_{k_1} < 1 - h'$. As a result, $\Delta_{k_1} \in (0,1)$, and the vector $\left(\frac{k_1 h}{K}, \Delta_{k_1}, ..., \Delta_{k_1}\right)$ is an element of $(0,1)^\Omega$. Therefore, we can rewrite (41) as

$$\hat{I}_{c_1'}\left(\frac{k_1 h}{K}, \Delta_{k_1}, ..., \Delta_{k_1}\right) = \hat{I}_{c_1'}\left(\frac{(k_1+1)h}{K}, 0, ..., 0\right).$$

Moreover, since $\Delta_{k_1} < 1 - h'$, it holds that $\Delta_{k_1} + h' \in (0,1)$. Therefore, the vector

$$\left(\frac{k_1 h}{K} + h', \Delta_{k_1} + h', ..., \Delta_{k_1} + h'\right)$$

is an element of $(0,1)^\Omega$.[39] As a result of (28) in Lemma A.6, it holds that

$$\hat{I}_{c_1'}\left(\frac{k_1 h}{K} + h', \Delta_{k_1} + h', ..., \Delta_{k_1} + h'\right) = \hat{I}_{c_1'}\left(\frac{(k_1+1)h}{K} + h', h', ..., h'\right).$$

Substitute the definition of $\hat{I}_{c_1'}$ into the equation above, then we have

$$\begin{aligned} &\Phi\left(\frac{k_1}{K}h + h'\right) + \sum_{\omega \neq \omega^*} \mu_{1\omega} \cdot f_{c_1}(\omega) \cdot \phi_{\theta(c_1)}(\Delta_{k_1} + h') \\ &\qquad\qquad = \Phi\left(\frac{k_1+1}{K}h + h'\right) + \sum_{\omega \neq \omega^*} \mu_{1\omega} \cdot f_{c_1}(\omega) \cdot \phi_{\theta(c_1)}(h'). \end{aligned} \tag{42}$$

Equations (41) and (42) can be rewritten as the following two equations respectively:

$$\Phi\left(\frac{k_1}{K}h\right) - \Phi\left(\frac{k_1+1}{K}h\right) = \sum_{\omega \neq \omega^*} \mu_{1\omega} \cdot f_{c_1}(\omega) \cdot [\phi_{\theta(c_1)}(\Delta_{k_1}) - \phi_{\theta(c_1)}(0)], \tag{43}$$

$$\Phi\left(\frac{k_1}{K}h + h'\right) - \Phi\left(\frac{k_1+1}{K}h + h'\right) = \sum_{\omega \neq \omega^*} \mu_{1\omega} \cdot f_{c_1}(\omega) \cdot [\phi_{\theta(c_1)}(\Delta_{k_1} + h') - \phi_{\theta(c_1)}(h')]. \tag{44}$$

Divide both sides of (44) by the respective sides in (43), we have

$$\frac{\Phi\left(\frac{k_1}{K}h + h'\right) - \Phi\left(\frac{k_1+1}{K}h + h'\right)}{\Phi\left(\frac{k_1}{K}h\right) - \Phi\left(\frac{k_1+1}{K}h\right)} = \frac{\phi_{\theta(c_1)}(\Delta_{k_1} + h') - \phi_{\theta(c_1)}(h')}{\phi_{\theta(c_1)}(\Delta_{k_1}) - \phi_{\theta(c_1)}(0)} = \exp[\theta(c_1)h'].$$

[39] The first entry $k_1 h/K + h' < h + h' < 1$, therefore it also lies in $(0,1)$.

This equation holds for all $k_1 = 0, 1, ..., K-1$. Add the numerators (*resp.* denominators) in the equations corresponding to $k_1 = 0, 1, ..., K-1$ together to form the numerator (*resp.* denominator) in (39), and we get exactly (39):

$$\frac{\Phi(h+h') - \Phi(h')}{\Phi(h) - \Phi(0)} = \exp\left[\theta(c_1)h'\right].$$

Rearrange this equation leads to $\Phi(h+h') = \exp[\theta(c_1)h'][\Phi(h) - \Phi(0)] + \Phi(h')$, which holds for all $h, h' \in (0,1)$ such that $(h+h') \in (0,1)$. Define the set $R \equiv \{(h,h') \mid h, h' > 0, h+h' < 1\}$, and $\mathcal{K}(h+h') \equiv \Phi(h+h')$, $\mathcal{M}(h') \equiv \exp[\theta(c_1)h']$, $\mathcal{N}(h) \equiv \Phi(h) - \Phi(0)$, and $\mathcal{L}(h') \equiv \Phi(h')$, then we get a functional equation in the form of $\mathcal{K}(h+h') = \mathcal{M}(h')\mathcal{N}(h) + \mathcal{L}(h')$, which by the corollary in Aczél (2005) implies that there exists $\hat{\theta} > 0$ and $\hat{\theta}, \hat{k}, \hat{b}, \hat{\hat{b}}, \tilde{b} \in \mathbb{R}$, such that either of the following two cases hold:

- *Case 1*: $\Phi(h) = \mathcal{K}(h) = \hat{\theta}h + \hat{b}$ and $\exp[\theta(c_1)h] = \mathcal{M}(h) = \tilde{b}$, or
- *Case 2*: $\Phi(h) = \mathcal{K}(h) = \hat{k}\exp(\hat{\theta}h) + \hat{\hat{b}}$ and $\exp[\theta(c_1)h] = \mathcal{M}(h) = \tilde{b}\exp\hat{\theta}h$.

In both cases, $\Phi(h) = k'\phi_{\theta(c_1)} + b'$ for some $k' > 0$ and $b' \in \mathbb{R}$. Specifically, $\theta(c_1) = 0$ in Case 1 and $\theta(c_1) > 0$ in Case 2. Substitute this back to (37), and we have $\trianglerighteq_{c_1'}$ is represented by

$$\hat{I}_{c_1'}(\zeta) = k'\phi_{\theta(c_1)}(\zeta_{\omega^*}) + \sum_{\omega \neq \omega^*} k\mu_{1\omega} \cdot f_{c_1}(\omega) \cdot \phi_{\theta(c_1)}(\zeta_\omega) + b + b'.$$

Now define $\theta(c_1') = \theta(c_1)$ and $f_{c_1'} : \Omega \to \mathbb{R}$ by

$$f_{c_1'}(\omega) = \begin{cases} \frac{k'}{\mu_{1\omega^*}} & \text{if } \omega = \omega^*; \\ k f_{c_1}(\omega) & \text{if } \omega \neq \omega^*. \end{cases}$$

It follows that $\hat{I}_{c_1'}(\zeta)$ is ordinally equivalent to

$$\sum_{\omega \in \Omega} \mu_{1\omega} \cdot f_{c_1'}(\omega) \cdot \phi_{\theta(c_1')}(\zeta_\omega),$$

that is, $\mathbb{E}_{\mu_1}[f_{c_1'}(\omega) \cdot \phi_{\theta(c_1')}(\zeta)]$. As a result, $\trianglerighteq_{c_1'}$ is represented by $\mathbb{E}_{\mu_1}[f_{c_1'}(\omega) \cdot \phi_{\theta(c_1')}(\zeta)]$ over $(0,1)^\Omega$. By the continuity of $\trianglerighteq_{c_1'}$, this result extends to $[0,1]^\Omega$.

$\square$

LEMMA A.11. *For every $c_1 \in \mathcal{C}_1$, there exists $f_{c_1} : \Omega \to \mathbb{R}_{++}$ and $\theta(c_1) \geq 0$ such that $I_{c_1}(\zeta)$ is ordinally equivalent to*

$$\mathbb{E}_{\mu_1}[f_{c_1}(\omega) \cdot \phi_{\theta(c_1)}(\zeta)]. \tag{45}$$

We call $f_{c_1} : \Omega \to \mathbb{R}_{++}$ the reweight function corresponding to consumption c_1. Moreover, there exists $\theta^* \geq 0$, such that $\theta(c_1) = \theta^*$ for all $c_1 \in \mathcal{C}_1$.

PROOF. Take an arbitrary $x \in X$. By Lemma A.9, $\trianglerighteq_x$ is represented by $\sum_\omega \mu_{1\omega} \exp(\theta_x w_\omega)$. Enumerate the states as $\omega_1, \omega_2, ..., \omega_n$ and for each n, define $E_n = \{\omega_n\}$. Fix an arbitrary $c_1^* \in \mathcal{C}_1$. Denote $c_1^{(0)} = x$ and $c_1^{(i+1)} = c_1^* E_n c_1^{(i)}$ for every $i = 0, 1, ..., n-1$. By definition, $c_1^{(n)} = c_1^*$. By Lemma A.10 and mathematical induction, there exists $f_{c_1^{(n)}} : \Omega \to \mathbb{R}_{++}$ such that $\trianglerighteq_{c_1^{(n)}}$ is represented by $\mathbb{E}_{\mu_1}\left[f_{c_1^{(n)}}(\omega) \cdot \phi_{\theta\left(c_1^{(n)}\right)}(\zeta)\right]$. That is, $\trianglerighteq_{c_1^*}$ is represented by $\mathbb{E}_{\mu_1}\left[f_{c_1^*}(\omega) \cdot \phi_{\theta(c_1^*)}(\zeta)\right]$. Moreover, mathematical induction implies that $\theta\left(c_1^*\right) = \theta(x)$. Since $c_1^* \in \mathcal{C}_1$ and $x \in X$ are arbitrarily chosen, there exists a constant $\theta^* \geq 0$ such that $\theta(c_1) = \theta$ for all $c_1 \in \mathcal{C}_1$. Finally, notice that if $f_{c_1}, g_{c_1} : \Omega \to \mathbb{R}$ satisfy that $\mathbb{E}_{\mu_1}[f_{c_1}(\omega) \cdot \phi_{\theta(c_1)}(\zeta)]$ and $\mathbb{E}_{\mu_1}[g_{c_1}(\omega) \cdot \phi_{\theta(c_1)}(\zeta)]$ are ordinally equivalent, then $f_{c_1} = k g_{c_1}$ for some $k > 0$. As a result, $f_{c_1} : \Omega \to \mathbb{R}$ in (45) is unique up to linear transformation. $\square$

For $E \in \mathcal{E}$ and $f, g : E \to \mathbb{R}$, we say $f \approx g$ if $f = kg$ for some $k > 0$. For $\mu \in Int(\Delta\Omega)$, let μ_1^E define the Bayesian posterior. For $\hat{f} : \Omega \to \mathbb{R}$, let $\hat{f}^E$ define the restriction of the function $\hat{f}$ within the domain E.

LEMMA A.12. *for any $E \in \mathcal{E}$, $f, g : \Omega \to \mathbb{R}$, $\theta \geq 0$, if $\mathbb{E}_{\mu_1^E}[f(\omega) \cdot \phi_\theta(\zeta)]$ and $\mathbb{E}_{\mu_1^E}[g(\omega) \cdot \phi_\theta(\zeta)]$ are ordinally equivalent, then*

$$f^E \approx g^E.$$

PROOF. By the uniqueness of expected utility representation, for each $\omega \in E$, it holds that

$$\frac{\mu_{1\omega} \cdot f(\omega)}{\sum_{\omega' \in E} \mu_{1\omega'} \cdot f(\omega')} = \frac{\mu_{1\omega} \cdot g(\omega)}{\sum_{\omega' \in E} \mu_{1\omega'} \cdot g(\omega')}.$$

Take

$$\hat{k} = \frac{\sum_{\omega' \in E} \mu_{1\omega'} \cdot f(\omega')}{\sum_{\omega' \in E} \mu_{1\omega'} \cdot g(\omega')},$$

then for each $\omega \in E$, it holds that $f(\omega) = \hat{k}g(\omega)$. Therefore, $f^E \approx g^E$. $\square$

LEMMA A.13. *For every $x, y \in X$ such that $|u(x) - u(y)| < \delta$, $E \in \mathcal{E}$, $\hat{c}_1 = xEy$ and $\omega \in \Omega$, then for reweight function $f_{\hat{c}_1}$, it holds that $f_{\hat{c}_1} \approx f^*_{\hat{c}_1}$, where*[40]

$$f^*_{\hat{c}_1}(\omega) = \begin{cases} \phi_{\theta^*/\delta}[u(x)] & \text{if } \omega \in E; \\ \phi_{\theta^*/\delta}[u(y)] & \text{if } \omega \notin E. \end{cases}$$

PROOF. Fix $x, y \in X$ such that $|u(x) - u(y)| < \delta$, $E \in \mathcal{E}$ and $\hat{c}_1 = xEy$. Fix $f_{\hat{c}_1}$ (*resp.* f_x, f_y) $: \Omega \to \mathbb{R}$ such that $\trianglerighteq_{\hat{c}_1}$ (*resp.* $\trianglerighteq_x, \trianglerighteq_y$) is represented by $\mathbb{E}_{\mu_1^E}[f_{\hat{c}_1}(\omega) \cdot \phi_{\theta^*}(\zeta)]$ (*resp.* $\mathbb{E}_{\mu_1^E}[f_x(\omega) \cdot \phi_{\theta^*}(\zeta)]$, $\mathbb{E}_{\mu_1^E}[f_y(\omega) \cdot \phi_{\theta^*}(\zeta)]$). First, notice that by Lemma A.6 (ii), $\trianglerighteq^E_{\hat{c}_1} = \trianglerighteq^E_x$. By Lemma A.9, $f_x \approx 1_\Omega$, which implies $f^E_x \approx 1_E$. By Lemma A.12, $f^E_{\hat{c}_1} \approx f^E_x \approx 1_E$. Similarly, $f^{E^c}_{\hat{c}_1} \approx f^{E^c}_y \approx 1_{E^c}$.

For reweight function $f_{\hat{c}_1}$, define $f_{\hat{c}_1}(E) \equiv \sum_{\omega \in E} f_{\hat{c}_1}(\omega)$ and $f_{\hat{c}_1}(E^c) \equiv \sum_{\omega \in E} f_{\hat{c}_1}(\omega)$. To prove the lemma, it suffices to show that

$$\frac{f_{\hat{c}_1}(E)}{f_{\hat{c}_1}(E^c)} = \frac{\phi_{\theta^*/\delta}[u(x)]}{\phi_{\theta^*/\delta}[u(y)]}. \tag{46}$$

Without loss of generality, assume $u(x) > u(y)$. Take $x', y' \in X$ such that $u(x') - u(y') = -[u(x) - u(y)]/\delta$; that is, $u(x) + \delta u(x') = u(y) + \delta u(y')$. Take $\hat{c} = (\hat{c}_1, \hat{c}_2)$, where $\hat{c}_1 = xEy$ and $\hat{c}_2 = x'Ey'$. For any $\tilde{c}_2 = \tilde{x}E\tilde{y}$ such that

$$\mathbb{E}_{\mu_1}[u(\tilde{c}_2)] > \mathbb{E}_{\mu_1}[u(\hat{c}_2)],$$

I claim that

$$u(\tilde{c}_2) \trianglerighteq_{\hat{c}_1} u(\hat{c}_2). \tag{47}$$

Without loss of generality, I assume that $u(\tilde{x}) > u(x')$. If $u(\tilde{x}) > u(x')$ and $u(\tilde{y}) > u(y')$, then (47) holds since $\trianglerighteq_{\hat{c}_1}$ is monotone. Therefore, I focus on the case where $u(\tilde{x}) > u(x')$ and $u(\tilde{y}) \leq u(y')$. To see why (47) is true in this case, consider the period-1 act $\hat{a} = (\hat{c}_1, \{\hat{c}_2, \tilde{c}_2\})$. Event E is desirable for $\hat{a}$ because

$$u(\hat{c}_{1E}) + \delta u(\hat{c}_{2E}) = u(x) + \delta u(x')$$

[40] In the expression below, the parameter $\theta^* \geq 0$ is defined in Lemma A.11.

$$= u(y) + \delta u(y') = u(\hat{c}_{1E^c}) + \delta u(\hat{c}_{2E^c}),$$

and

$$\begin{aligned} u(\hat{c}_{1E}) + \delta u(\tilde{c}_{2E}) &= u(x) + \delta u(\tilde{x}) \\ &> u(x) + \delta u(x') \\ &= u(y) + \delta u(y') \\ &\geq u(y) + \delta u(\tilde{y}) = u(\hat{c}_{1E^c}) + \delta u(\tilde{c}_{2E^c}). \end{aligned}$$

Moreover, the *ex-ante* self would choose $\tilde{c}_2$ over $\hat{c}_2$ because $\mathbb{E}_{\mu_1}[u(\tilde{c}_2)] > \mathbb{E}_{\mu_1}[u(\hat{c}_2)]$. Conditional on event E, the payoff of $\tilde{c}_2$ is still higher than $\hat{c}_2$ due to the assumption that $u(\tilde{x}) > u(x')$.[41] Therefore, by Axiom 4 (Cognitive Consistency), it holds that $\gamma(\hat{a}) = (\hat{c}_1, \tilde{c}_2)$. That is, $\tilde{c}_2 \succsim^*_{\hat{c}_1} \hat{c}_2$, ***i.e.*** $u(\tilde{c}_2) \trianglerighteq_{\hat{c}_1} u(\hat{c}_2)$. Therefore, if

$$\mu_{1E} u(\tilde{x}) + \mu_{1E^c} u(\tilde{y}) > \mu_{1E} u(x') + \mu_{1E^c} u(y'),$$

then

$$\mu_{1E} \cdot f_{\hat{c}_1}(E) \cdot \phi_{\theta^*}[u(\tilde{x})] + \mu_{1E^c} \cdot f_{\hat{c}_1}(E^c) \cdot \phi_{\theta^*}[u(\tilde{y})] \geq \mu_{1E} \cdot f_{\hat{c}_1}(E) \cdot \phi_{\theta^*}[u(x')] + \mu_{1E^c} \cdot f_{\hat{c}_1}(E^c) \cdot \phi_{\theta^*}[u(y')].$$

Therefore, for $h, h' \in [0,1]$,

$$\begin{aligned} \mu_{1E} h + \mu_{1E^c} h' &= \mu_{1E} u(x') + \mu_{1E^c} u(y') \\ \implies \quad & \mu_{1E} \cdot f_{\hat{c}_1}(E) \cdot \phi_{\theta^*}(h) + \mu_{1E^c} \cdot f_{\hat{c}_1}(E^c) \cdot \phi_{\theta^*}(h') \\ & \geq \mu_{1E} \cdot f_{\hat{c}_1}(E) \cdot \phi_{\theta^*}[u(x')] + \mu_{1E^c} \cdot f_{\hat{c}_1}(E^c) \cdot \phi_{\theta^*}[u(y')]. \end{aligned}$$

As a result, at $(h, h') = [u(x'), u(y')]$, the following statement holds:

$$\begin{aligned} \mu_{1E} dh + \mu_{1E^c} dh' &= 0 \\ \implies \quad & \mu_{1E} \cdot f_{\hat{c}_1}(E) \cdot d\phi_{\theta^*}(h) + \mu_{1E^c} \cdot f_{\hat{c}_1}(E^c) \cdot d\phi_{\theta^*}(h') = 0, \end{aligned}$$

[41] Recall that when constructing $\tilde{c}_2 = \tilde{x} E \tilde{y}$, I assume that $\tilde{c}_2$ does not dominate $\hat{c}_2$ in all states. Moreover, I assumed $u(\tilde{x}) > u(x')$ and $u(\tilde{y}) \leq u(y')$. This assumption is without loss of generality because, if otherwise $u(\tilde{x}) \leq u(x')$ and $u(\tilde{y}) > u(y')$, then E^c, instead of E, will be desirable for $\hat{a}$. All other parts of the proof follow as before.

which implies

$$\frac{f_{\hat{c}_1}(E)}{f_{\hat{c}_1}(E^c)} = \frac{\phi_{\theta^*}[u(y')]}{\phi_{\theta^*}[u(x')]}$$
$$= \phi_{\theta^*}[u(y') - u(x')]$$
$$= \phi_{\theta^*}\left[\frac{u(x) - u(y)}{\delta}\right] = \phi_{\theta^*/\delta}[u(x) - u(y)],$$

as desired. □

LEMMA A.14. *For every $c_1 \in \mathcal{C}_1$, $\psi, \psi' \in \Omega$, and reweight function f_{c_1}, it holds that*

$$\frac{f_{c_1}(\psi)}{f_{c_1}(\psi')} = \frac{\phi_{\theta^*/\delta}[u(c_{1\psi})]}{\phi_{\theta^*/\delta}[u(c_{1\psi'})]} \tag{48}$$

PROOF. Fix $c_1 \in \mathcal{C}_1$ and $\psi, \psi' \in \Omega$. Define $\hat{E} = \{\psi, \psi'\}$. First, suppose $|u(c_{1\psi}) - u(c_{1\psi'})| < \delta$, and take $c_1' \in \mathcal{C}_1$ such that $\hat{c}_{1\psi} = c_{1\psi}$ and $\hat{c}_{1\omega} = c_{1\psi'}$ for all $\omega \neq \psi$. By Lemma A.13, for reweight function $f_{c_1'} : \Omega \to \mathbb{R}_{++}$, it holds that

$$\frac{f_{c_1'}(\psi)}{f_{c_1'}(\psi')} = \frac{\phi_{\theta^*/\delta}[u(c_{1\psi})]}{\phi_{\theta^*/\delta}[u(c_{1\psi'})]}. \tag{49}$$

By Lemma A.6 (ii), $\trianglerighteq_{c_1}^{\hat{E}} = \trianglerighteq_{c_1'}^{\hat{E}}$. Therefore, Lemma A.12 implies that $f_{c_1}^{\hat{E}} \approx f_{c_1'}^{\hat{E}}$, which, together with (49), implies (48).

Next, I extend this result to the case where $|u(c_{1\psi}) - u(c_{1\psi'})| \geq \delta$ using an inductive argument. Hypothesize that for some $B \in (0,1)$, (48) holds for all $\hat{c}_1 \in \mathcal{C}_1$ such that $|u(\hat{c}_{1\psi}) - u(\hat{c}_{1\psi'})| < B$. I want to show that (48) holds for all $\tilde{c}_1 \in \mathcal{C}_1$ such that $|u(\tilde{c}_{1\psi}) - u(\tilde{c}_{1\psi'})| < 2B$. Fix such a $\tilde{c}_1$. Consider a state of the world $\psi'' \in \Omega - \{\psi, \psi'\}$, and I construct $\hat{\hat{c}}_1$ such that $\hat{\hat{c}}_{1\omega} = \tilde{c}_{1\omega}$ for $\omega \in \{\psi, \psi'\}$ and $\hat{\hat{c}}_{1\psi''} = \frac{1}{2}\tilde{c}_{1\psi} + \frac{1}{2}\tilde{c}_{1\psi'}$. As a result,

$$|u(\hat{\hat{c}}_\psi) - u(\hat{\hat{c}}_{\psi''})| = |u(\hat{\hat{c}}_{\psi'}) - u(\hat{\hat{c}}_{\psi''})| = \frac{1}{2}|u(\hat{\hat{c}}_\psi) - u(\hat{\hat{c}}_{\psi'})| < B.$$

By the inductive hypothesis, we can apply (48) to the following two pairs of states: $\{\psi, \psi''\}$, and $\{\psi', \psi''\}$, which implies

$$\frac{f_{\hat{\hat{c}}_1}(\psi)}{f_{\hat{\hat{c}}_1}(\psi'')} = \frac{\phi_{\theta^*/\delta}[u(\hat{\hat{c}}_{1\psi})]}{\phi_{\theta^*/\delta}[u(\hat{\hat{c}}_{1\psi''})]}$$

and

$$\frac{f_{\hat{c}_1}(\psi'')}{f_{\hat{c}_1}(\psi')} = \frac{\phi_{\theta^*/\delta}[u(\hat{\hat{c}}_{1\psi''})]}{\phi_{\theta^*/\delta}[u(\hat{\hat{c}}_{1\psi'})]}$$

Multiply both sides of the first equation by the corresponding side of the second, and we have

$$\frac{f_{\hat{c}_1}(\psi)}{f_{\hat{c}_1}(\psi')} = \frac{\phi_{\theta^*/\delta}[u(\hat{\hat{c}}_{1\psi})]}{\phi_{\theta^*/\delta}[u(\hat{\hat{c}}_{1\psi'})]}.$$

Since for $\hat{E} = \{\psi, \psi'\}$, $\hat{\hat{c}}_1^{\hat{E}} = \tilde{c}_1^{\hat{E}}$, by Lemma A.12, it holds that $f^E_{\tilde{c}_1} \approx f^E_{\hat{\hat{c}}_1}$, which implies

$$\frac{f_{\tilde{c}_1}(\psi)}{f_{\tilde{c}_1}(\psi')} = \frac{\phi_{\theta^*/\delta}[u(\tilde{c}_{1\psi})]}{\phi_{\theta^*/\delta}[u(\tilde{c}_{1\psi'})]},$$

as desired. Recall that in the first half of this proof, I showed that (48) holds if $|u(c_{1\psi}) - u(c_{1\psi'})| < \delta$. According to the inductive argument above, (48) holds if $|u(c_{1\psi}) - u(c_{1\psi'})| < 2\delta$, 4δ,... Since $\delta > 0$, (48) holds for all $c_1 \in \mathcal{C}_1$. □

Let $\theta = \theta^*/\delta$. By Lemma A.14, for every $c_1 \in \mathcal{C}_1$, there exists some $k > 0$ such that for each $\omega \in \Omega$, the reweight function satisfies $f_{c_1}(\omega) = k\phi_\theta[u(c_{1\omega})]$. Without loss of generality, assume $k = 1$. By Lemma A.11, the order $\trianglerighteq_{c_1}$ is represented by $\mathbb{E}_{\mu_1}[f_{c_1}(\omega) \cdot \phi_{\delta\theta}(\zeta)]$. Substitute the closed-form of $f_{c_1}(\cdot)$ into the representation, $\trianglerighteq_{c_1}$ is represented by $\mathbb{E}_{\mu_1}[\phi_\theta(u(c_1)) \cdot \phi_{\delta\theta}(\zeta)]$, which equals $\mathbb{E}_{\mu_1}[\phi_\theta(u(c_1) + \delta\zeta)]$. Therefore, for any finite $C_2 \subset \mathcal{C}_2$, $(c_1, \hat{c}_2) \in \gamma_2[(c_1, C_2)]$ if and only if

$$\hat{c}_2 \in \arg\max_{c_2 \in C_2} \mathbb{E}_{\mu_1}[\phi_\theta(u(c_1) + \delta u(c_2))]. \tag{50}$$

For any period-1 action a, by Proposition 3.1, (50) implies that $\gamma(a) = \gamma_\kappa(a)$ for $\kappa = (\mu_1, u, \delta, \theta) \in K^*$. Moreover, for any 1-determined menu where the DM has full commitment power, Theorem A.1 implies that $\gamma(A) = \gamma_\kappa(A)$ for the same parameter κ as defined above. Now, it remains to show that $\gamma(A) = \gamma_\kappa(A)$ for any generic menu A.

LEMMA A.15. *Consider $\hat{A} = \{a^i\}_{i=1}^n$, where $a^i = (c_1^i, C_2^i)$ for every $i \in I \equiv \{1, 2, ..., n\}$. It holds that*

$$\gamma(\hat{A}) = \gamma_\kappa(\hat{A}).$$

PROOF. I first show that $\gamma(A) \subset \gamma_\kappa(A)$. Fix $c^* = (c_1^*, c_2^*) \in \gamma(\hat{A})$. By Axiom 3 (Sophistication), there exists a consistent selection $\hat{A}' = \{c^i\}_{i \in I}$ from $\hat{A}$ and $j \in I$, such that $c^* = c^j \in \gamma(\hat{A}')$.

It suffices to show that $a^j = (c_1^j, C_2^j)$ and $c_2^j \in C_2^j$ are the equilibrium period-1 and period-2 choices in a SPNE of the CD game with menu $\hat{A}$ and parameter κ.

First, since $\hat{A}' = \{c^i\}_{i \in I}$ is a consistent selection, for each $i \in I$, $c^i \in \gamma(a^i)$. Recall that (50) above establishes that $\gamma(a) = \gamma_\kappa(a)$ for every $a \in \mathbb{A}$. Therefore, for each $i \in I$, it holds that $c^i \in \gamma_\kappa(a^i)$. That is, for each period-1 choice $a^i \in \hat{A}$, the corresponding subgame has c^i as an equilibrium period-2 consumption.

Now, consider the backward induction problem for the period-1 self. She correctly anticipates that, for each $i \in I$, her future self will choose c_2^i if she chooses a^i. In this case, it is optimal for her to choose a^j, which leads to the consumption stream c^j. The choice a^j is indeed optimal for the period-1 self, because by Theorem A.1, $c^j \in \gamma(\hat{A}')$ implies $j \in \arg\max_{j' \in I} U_\mu(c^{j'})$. Through backward induction, we establsh that $a^j = (c_1^j, C_2^j)$ and $c_2^j \in C_2^j$ are equilibrium period-1 and period-2 choices in a SPNE of the CD game with menu $\hat{A}$ and parameter κ. Therefore, $c^* \in \gamma_\kappa(\hat{A})$. As a result, $\gamma(\hat{A}) \subset \gamma_\kappa(\hat{A})$.

Now I show that $\gamma(\hat{A}) \supset \gamma_\kappa(\hat{A})$. Suppose $c^* \in \gamma_\kappa(\hat{A})$, then there exists a SPNE such that (1) the period-1 self chooses $a^k = (c_1^k, C_2^k)$ (where $k \in I$), (2) for each $i \in I$, the period-2 self chooses $\hat{c}_2^i$ for the subgame corresponding to a^i, and (3) $c_1^k = c_1^*$ and $c_2^k = \hat{c}_2^*$. Then consider the set $\hat{\hat{A}} = \{(c_1^i, \hat{c}_2^i)\}$. For each $i \in I$, since $\hat{c}_2^i$ is the period-2 equilibrium choice, it holds that $(c_1^i, \hat{c}_2^i) \in \gamma_\kappa(a^i) = \gamma(a^i)$. Therefore, $\hat{\hat{A}}$ is a consistent selection from $\hat{A}$. Moreover, $(c_1^k, \hat{c}_2^k) \in \gamma_\kappa(\hat{\hat{A}}) = \gamma(\hat{\hat{A}})$ since the period-1 self chooses a^k in the SPNE. As a result, it holds that $c^* \in \gamma(\hat{A})$, the desired result. $\square$

A.4 *Proof of Theorem 3.1*

I now prove the "only if" direction of Theorem 3.1, as the "if" direction follows directly. Suppose Axioms 1, 2, 3 and DC hold for an induced choice correspondence γ. Axiom DC states that, for period-1 action a, $\gamma[F(a)] = c^* \implies \gamma(a) = c^*$. As a result, for any $c_1 \in \mathcal{C}_1$ and $c_2, c_2' \in \mathcal{C}_2$, it holds that $c_2 \succ^* c_2' \implies c_2 \succ_{c_1}^* c_2'$. Since $\succsim^*$ and $\succsim_{c_2}^*$ are continuous weak orders, it holds that $c_2 \succsim^* c_2' \iff c_2 \succsim_{c_1}^* c_2'$. Recall that according to Theorem A.1, Axiom 1 and 2 imply that $\succsim^*$ has a time-separable expected utility representation $\mathbb{E}_{\mu_1}[u(c_1)] + \delta \mathbb{E}_{\mu_1}[u(c_2)]$. As a result, $\succsim_{c_1}^*$ is represented by $\mathbb{E}_{\mu_1}[u(c_2)]$, which implies that Axioms 5 and 6 hold. Since Axioms 1-6 hold, Theorem 3.2 implies the existence of $\kappa = (\mu_1, u, \delta, \theta)$ such that $\gamma = \gamma_\kappa$.

Now it remains to prove that $\theta = 0$. Suppose for a contradiction that $\theta > 0$. Take $\hat{c}_2 \in \mathcal{C}_2$ and $\hat{x} \in X$, such that (1) $u \circ \hat{c}_2 \in \mathbb{R}^\Omega$ is not constant across all states, and (2) $\mathbb{E}_{\mu_1}[u(\hat{c}_2)] = u(\hat{x})$. Since $\theta > 0$ and μ_1 has full support, it holds that $\mathbb{E}_{\mu_1}[\phi_{\delta\theta} \circ u(\hat{c}_2)] > \phi_{\delta\theta} \circ u(\hat{x})$. Therefore,

$\gamma(\{(\hat{x},\hat{c_2}),(\hat{x},\hat{x})\}) = \{(\hat{x},\hat{c_2}),(\hat{x},\hat{x})\}$ while $\gamma(\{(\hat{x},\{\hat{c_2},\hat{x}\})\}) = \{(\hat{x},\hat{c_2})\}$, which implies Axiom DC is violated. As a result, $\theta = 0$, which finishes the proof.

A.5 *Proof of Theorem 3.3*

I focus on the "only if" direction. Suppose for $\kappa = (\mu_1, u, \delta, \theta)$ and $\kappa' = (\mu_1', u, \delta', \theta')$, it holds that γ_κ is more CD-susceptible than $\gamma_{\kappa'}$. First, by definition, $\gamma_\kappa(A) = \gamma_\kappa'(A)$ for all $A \in \mathcal{A}_1$. As a result, with full commitment power *ex-ante*, the choices according to γ_κ and $\gamma_{\kappa'}$ are completely identical. By the uniqueness property of the SEU representation, $\mu_1 = \mu_1'$ and $\delta = \delta'$. Now suppose for a contradiction that $\theta < \theta'$. in that case, take a non-empty proper subset E of Ω. Take $\hat{x} \in X$ and $\hat{c}_2 \in \mathcal{C}_2$ such that (1) $\hat{c}_2$ is measurable with respect to $\{E, E^c\}$, (2) $u(\hat{c}_{2E}) > u(\hat{c}_{2E^c})$, and (3) $\mathbb{E}_{\mu_1}[\phi_{\delta\theta'} \circ u(\hat{c}_2)] = \phi_{\delta\theta'} \circ u(\hat{x})$. Since $\theta < \theta'$, it holds that $\mathbb{E}_{\mu_1}[\phi_{\delta\theta} \circ u(\hat{c}_2)] < \phi_{\delta\theta} \circ u(\hat{x})$ and $\mathbb{E}_{\mu_1}[u(\hat{c}_2)] < u(\hat{x})$. As a result, for $\hat{a} = (\hat{x}, \{\hat{x}, \hat{c}_2\})$, it holds that $\gamma_\kappa[F(\hat{a})] = \gamma_{\kappa'}[F(\hat{a})] = (\hat{x},\hat{x})$, $\gamma_\kappa(\hat{a}) = (\hat{x},\hat{x})$ and $\gamma_{\kappa'}(\hat{a}) = \{(\hat{x},\hat{x}),(\hat{x},\hat{c}_2)\}$. Therefore, γ_κ is $\hat{a}$-DC while $\gamma_{\kappa'}$ is not. Therefore, γ_κ cannot be more CD-susceptible than $\gamma_{\kappa'}$, a contradiction. As a result, it must hold that $\theta \geq \theta'$, which finishes the proof.

Appendix B: Proof of Other Results

In this section, I introduce the proof of the results in Section 4 and 5 of this paper.

B.1 *Proof of Results in Section 4*

I introduce the proof for the results in Section 4.3, which apply the CD game to real-life phenomena such as effort justification, induced compliance, the negative incentive effect, and polarization.

B.1.1 *Proof of Proposition 4.1* I start the proof by introducing the closed-form expression for the equilibrium period-2 belief in the CD game.

Lemma B.1. *In the CD game with parameter $\kappa = (\mu_1, u, \delta, \theta) \in K$ and equilibrium consumption stream c^*, the equilibrium period-2 belief is given by $\mu_2^* = (\mu_{2\omega}^*)_{\omega\in\Omega}$ such that for every $\omega \in \Omega$,*

$$\mu_{2\omega}^* = \frac{\mu_{1\omega}\exp[\theta u(c_{1\omega}^*) + \delta\theta u(c_{2\omega}^*)]}{\sum\limits_{\psi\in\Omega}\mu_{1\psi}\exp[\theta u(c_{1\psi}^*) + \delta\theta u(c_{2\psi}^*)]}.$$

PROOF. Recall that (16) in Lemma A.1 states that for equilibrium period-1 consumption c_1^*, the equilibrium period-2 belief μ_2^* and period-2 consumption c_2^* jointly solve

$$\max_{(\mu,c_2)\in\Delta\Omega\times C_2} \theta U_\mu(c_1^*, c_2) - D_{KL}(\mu||\mu_1).$$

As a result, for the equilibrium consumption stream $c^* = (c_1^*, c_2^*)$, the equilibrium belief μ_2^* must solves

$$\max_{\mu\in\Delta\Omega} \theta U_\mu(c_1^*, c_2^*) - D_{KL}(\mu||\mu_1).$$

The Lagrangian of this problem is given by

$$\mathcal{L}(\mu,\lambda) = \theta\sum_{\psi\in\Omega}\mu_\psi[u(c_{1\psi}) + \delta u(c_{2\psi})] - \sum_{\psi\in\Omega}\mu_\psi \log\left(\frac{\mu_\psi}{\mu_{1\psi}}\right) - \lambda\left[\sum_{\psi\in\Omega}\mu_\psi - 1\right].$$

Solve for the first-order condition, and for any $\omega,\omega'\in\Omega$,

$$\frac{\mu_\omega}{\mu_{\omega'}} = \frac{\mu_{1\omega}}{\mu_{1\omega'}}\frac{\exp\left[\theta\left(u(c_{1\omega}) + \delta u(c_{2\omega})\right)\right]}{\exp\left[\theta\left(u(c_{1\omega'}) + \delta u(c_{2\omega'})\right)\right]},$$

which implies the closed-form solution in Lemma B.1. □

Recall that $S = \{l, r\}$, where $l < 0$ and $r > 0$. For period-1 choice $s_1 \in S$ and pressure $\chi \in \mathcal{X}$, let $s_2^*(s_1,\chi) \in S$ define equilibrium investment choice at $t = 2$. Recall from Section 4.3 that $c(s,\chi)$ defines the consumption induced by choice s and social pressure χ. In the subgame where the DM chooses $s_1 \in S$ at $t = 1$, I define the DM's equilibrium payoff by $\mathbb{U}(s_1,\chi) = \mathbb{E}[u\circ c(s_1,\chi)] + \delta\mathbb{E}[u\circ c(s_2^*, 0)]$. In the expression, the equilibrium period-2 investment $s_2^*(s_1,\chi)$ is simplified as s_2^*.

LEMMA B.2. *It holds that*

1. $s_2^*(s_1,\chi)$ *is weakly increasing in* χ *for* $s_1 \in \{l, r\}$*;*
2. $\mathbb{U}(r,\chi)$ *is strictly decreasing in* χ*;*
3. $\mathbb{U}(l,\chi)$ *is strictly increasing in* χ*.*

PROOF. For $s = l, r$ and $\chi \geq 0$, I first define $\Delta U(s,\chi) \equiv u(s - \chi s) - u(-s - \chi s)$. Since $u'' > 0$ and $s \neq 0$, it holds that $\Delta U(s,\chi)$ is strictly increasing with χ. Moreover, it holds that $\Delta U(s,\chi)$ is strictly increasing in s. To facilitate my analysis, I define $v \equiv \phi_{\delta\theta}\circ u$, where u is the utility index and $\phi_{\delta\theta}$ is the transformation defined in (10). By the closed-form solution for period-2 choices in Proposition 3.1, in the subgame following any period-1 choice $s_1 \in \{l, r\}$ and pressure level χ, the period-2 self chooses $s_2^*(s_1, \chi)$ that solves

$$\max_{s\in\{l,r\}} \mu_{1R} \cdot \exp[\theta \Delta U(s_1, \chi)] \cdot v(s) + \mu_{1L} \cdot v(-s). \tag{51}$$

For $s_1 \in \{l, r\}$, it holds that $\exp[\theta \Delta U(s_1, \chi)]$ is increasing in χ. As a result, the solution to (51) also increases with respect to χ, which implies $s_2^*(s_1, \chi)$ weakly increases with χ.

Now, it remains to prove statements 2 and 3 of the lemma are true. First, I claim that when $s_1 = r$, the payoff $\mathbb{U}(s_1, \chi) = \mathbb{E}[u \circ c(s_1, \chi)] + \delta \mathbb{E}[u \circ c(s_2^*, 0)]$ is strictly decreasing in χ. First, the period-1 payoff $\mathbb{E}[u \circ c(s_1, \chi)]$ is strictly decreasing in χ as the external pressure χ punishes the choice $s_1 = r$. Second, the period-2 payoff $\delta \mathbb{E}[u \circ c(s_2^*, 0)]$ is weakly decreasing in χ because $s_2^*(r, 0) = r$. As a result, the corresponding period-2 payoff $\delta \mathbb{E}[u \circ c(s_2^*, 0)]$ has to be weakly decreasing in χ.

Finally, I claim that when $s_1 = l$, the payoff $\mathbb{U}$ is strictly increasing in χ. First, the period-1 payoff $\mathbb{E}[u \circ c(s_1, \chi)]$ is strictly increasing in χ as the external pressure χ rewards the choice $s_1 = l$. Second, the period-2 payoff $\delta \mathbb{E}[u \circ c(s_2^*, 0)]$ is weakly increasing in χ. To see why, notice that $s_2^*(l, \chi) \in \{l, r\}$ is weakly increasing in χ. As a result, the payoff given by $s_2^*(l, \chi)$, $\delta \mathbb{E}[u \circ c(s_2^*, 0)]$, is also weakly increasing in χ, because $r > l$ and r is more valuable than l *ex-ante*.

□

By Lemma B.2, $\mathbb{U}(r, \chi)$ is strictly decreasing in χ while $\mathbb{U}(l, \chi)$ is strictly increasing in χ. Moreover, $\mathbb{U}(r, 0) > \mathbb{U}(l, 0)$ by assumption and $\mathbb{U}(r, \chi) < \mathbb{U}(l, \chi)$ for sufficiently large initial pressure χ. As a result, there exists a unique χ^* such that the equilibrium period-1 investment $s_1^*(\chi) = r$ if $\chi < \chi^*$ and $s_1^*(\chi) = l$ if $\chi > \chi^*$.

Finally, recall that $\Delta U(r, \chi)$ is defined by $u(s - \chi r) - u(-s - \chi r)$, which is strictly increasing with s and χ. Within the interval $I = [0, \chi^*)$ or (χ^*, ∞), $s_1^*(\chi)$ is unchanged while $s_2^*(\chi)$ is weakly increasing. As a result, within interval I, the following two utility difference terms

$$\Delta_1(\chi) \equiv \Delta U(s_1^*(\chi), \chi)$$

and

$$\Delta_2(\chi) \equiv \Delta U(s_2^*(\chi), 0)$$

are both increasing in χ. The second term can be constant since $s_2^*(\chi)$ is only *weakly* increasing in χ. However, the first term $\Delta U(s_1^*(\chi),\chi)$ must be strictly increasing in χ. According to Lemma B.1,

$$\frac{\mu_{2R}^*(\chi)}{\mu_{2L}^*(\chi)} = \frac{\mu_{1R}}{\mu_{1L}} \cdot \exp[\theta\Delta_1(\chi) + \theta\delta\Delta_2(\chi)],$$

which implies that $\mu_{2R}^*(\chi)$ being strictly increasing in χ in $I = [0,\chi^*)$ or (χ^*,∞).

For all $\chi \in (0,\chi^*)$ and $\chi' \in (\chi^*,\infty)$, it holds that

$$s_1^*(\chi) = r > l = s_1^*(\chi');$$

moreover, $s_2^*(\chi) = r$ while $s_2^*(\chi') \in \{l,r\}$. Therefore, as $\chi \uparrow \chi^*$,

$$\lim_{\chi\uparrow\chi^*} \frac{\mu_{2R}^*(\chi)}{\mu_{2L}^*(\chi)} \geq \frac{\mu_{1R}}{\mu_{1L}} \cdot \exp[\theta\Delta U(r,\chi^*) + \theta\delta\Delta U(r,0)];$$

on the other hand, as $\chi \downarrow \chi^*$,

$$\lim_{\chi\uparrow\chi^*} \frac{\mu_{2R}^*(\chi)}{\mu_{2L}^*(\chi)} \leq \frac{\mu_{1R}}{\mu_{1L}} \cdot \exp[\theta\Delta U(l,\chi^*) + \theta\delta\Delta U(r,0)].$$

Since $\Delta U(r,\chi^*) > \Delta U(l,\chi^*)$, it holds that

$$\lim_{\chi\uparrow\chi^*} \frac{\mu_{2R}^*(\chi)}{\mu_{2L}^*(\chi)} > \lim_{\chi\uparrow\chi^*} \frac{\mu_{2R}^*(\chi)}{\mu_{2L}^*(\chi)},$$

the desired result.

B.1.2 *Proof of Proposition 4.2* Suppose $\chi = 0$, then the DM chooses r in both periods. According to the previous analysis,

$$\frac{\mu_{2R}^*(0)}{\mu_{2L}^*(0)} = \frac{\mu_{1R}}{\mu_{1L}} \cdot \exp[\theta\Delta U(r,0) + \theta\delta\Delta U(r,0)],$$

where $\Delta U(r,0)$ is strictly positive. Therefore, it holds that $\mu_{2+}^*(\chi) > \mu_{1+}$.

B.1.3 *Proof of Proposition 4.3* Recall that I assume the investor's period-1 asset choices equals Z_1. In this case, in the free-choice scenario, the DM's belief at $t=1$ is $\mu_{1,free}(Z_1,h) = \frac{2}{3}$

and $\mu_{1,free}(Z_2,l)=\frac{1}{3}$. By Lemma B.1,

$$\frac{P^*_{Z_1,free}(L)}{1-P_{Z_1,free}(L)}=\frac{\dfrac{2}{3}\cdot\exp(\theta\delta)}{\dfrac{1}{3}\cdot\exp(-\theta L)}=2\exp(\theta L+\theta\delta).$$

In the assigned-choice scenario, the DM's belief at $t=1$ is $\mu_{1,assigned}(Z_1,h)=\frac{4}{9}$, $\mu_{1,assigned}(Z_1,l)=\mu_{1,assigned}(Z_2,h)=\frac{2}{9}$, while $\mu_{1,assigned}(Z_2,l)=\frac{1}{9}$. By Lemma B.1,

$$\frac{P^*_{Z_1,assign}(L)}{1-P_{Z_1,assign}(L)}=\frac{\dfrac{4}{9}\cdot\exp(\theta\delta)+\dfrac{2}{9}}{\dfrac{2}{9}\cdot\exp(-\theta L+\theta\delta)\cdot\dfrac{1}{9}\cdot\exp(-\theta L)}<2\exp(\theta L+\theta\delta).$$

Therefore, for all $L>0$, it holds that $P^*_{Z_1,free}(L)>P^*_{Z_1,assign}(L)$. Moreover, as $L\to 0$,

$$\frac{P^*_{Z_1,free}(L)}{1-P_{Z_1,free}(L)}\to 2\exp(\theta\delta)>2,$$

while

$$\frac{P^*_{Z_1,assign}(L)}{1-P_{Z_1,assign}(L)}\to 2.$$

Therefore, $\lim_{L\to 0}P^*_{Z_1,free}(L)>\lim_{L\to 0}P^*_{Z_1,assign}(L)$.

B.2 *Proof of Results in Section 5.1*

In this proof, I assume that consumers break ties in favor of the seller. However, since the type distribution is assumed to be atomless, my analysis holds for any tie-breaking rule. Throughout this section, fix a dissonance parameter $\theta>0$. For any type π, the *ex-ante* willingness-to-buy for the add-on is given by $\alpha\pi$. Suppose the seller sets price P_1 for the basic product and P_2 for the add-on, and consider the subgame where type π purchases the basic product. Then by Proposition 3.1, the period-2 self is willing to purchase the add-on if and only if

$$\pi\phi_\theta(1+\alpha-P_1-P_2)+(1-\pi)\phi_\theta(-P_1-P_2)\geq\pi\phi_\theta(1-P_1)+(1-\pi)\phi_\theta(-P_1). \tag{52}$$

I let $v_\theta(\pi)$ define the willingness to pay for the add-on in $t=2$ conditional on the consumer π buying the basic product in $t=1$; that is, given the consumer buys the good in $t=1$, she would continue to buy the add-on if and only if $P_2\leq v_\theta(\pi)$. For the willingness to pay for the add-on, (52) implies

$$v_\theta(\pi) = \frac{1}{\theta} \ln \frac{[e^{(1+\alpha)\theta} - 1]\pi + 1}{(e^\theta - 1)\pi + 1}.$$

In a subgame where all customers in $[0,1]$ purchased the basic product, $v_\theta(\pi) \circ \hat{P}(Q) : [0,1] \to [0,\alpha]$ would be the inverse demand function for the add-on at time $t = 2$. Denote the corresponding price elasticity of demand by

$$\varepsilon_\theta(Q) \equiv -\frac{v_\theta \circ \hat{P}(Q)}{(v_\theta \circ \hat{P})'(Q) \cdot Q}.$$

LEMMA B.3. *For any $Q \in (0,1)$, it holds that*

$$\varepsilon_\theta(Q) > \varepsilon_0(Q).$$

PROOF. First, I show that v_θ is a strictly concave function. Consider the function $\hat{v}_\theta(\pi) \equiv \frac{[e^{(1+\alpha)\theta}-1]\pi+1}{(e^\theta-1)\pi+1}$, whose second-order derivative is strictly negative for $\pi \in [0,1]$, therefore, $\hat{v}_\theta$ is strictly concave. Since v_θ is a concave transformation of $\hat{v}_\theta$, it holds that v_θ is also strictly concave.

For v_θ and $P \in (0,1)$, denote $\varepsilon_{v_\theta}(P) \equiv \frac{v_\theta(P)}{v'_\theta(P) \cdot P}$. Notice that v_θ is strictly concave, with $v_\theta(0) = 0$ and $v_\theta(1) = \alpha$. As a result, it holds that $\varepsilon_{v_\theta}(P) > 1$ for every $P \in (0,1)$. Fix an arbitrary $Q \in (0,1)$; by the chain rule, the price elasticity of demand satisfies

$$\varepsilon_\theta(Q) = \varepsilon_{v_\theta}[\hat{P}(Q)] \cdot \varepsilon_0(Q),$$

which implies $\varepsilon_\theta(Q) > \varepsilon_0(Q)$. □

B.2.1 *Proof of Proposition 5.1*

DEFINITION B.1. *Denote $\mathbb{Q}_{\theta,2} \equiv \arg\max_{Q \in [0,1]}[Q \cdot (v_\theta \circ \hat{P}(Q))]$.*

Consider a hypothetical subgame where all consumers in the interval $[0,1]$ purchase the basic product. In this subgame, the $\mathbb{Q}_{\theta,2}$ defined above is the set of equilibrium quanitities of add-on sales. This auxiliary quantity plays a crucial role in the proof of Proposition 5.1.

LEMMA B.4. *For any $Q \in \mathbb{Q}_{\theta,2}$, it holds that $Q > Q^*_{0,2}$. Moreover, if $\pi \sim Uni[0,1]$, then $\mathbb{Q}_{\theta,2}$ is a singleton.*

PROOF. In Section 5.1, it is assumed that $Q^*_{0,2}$ is unique. As a result, $\varepsilon_0(Q) \geq 1$ for all $Q \leq Q^*_{0,2}$. By Lemma B.3, it holds that $\varepsilon_\theta(\hat{Q}) > 1$ for all $\hat{Q} \leq Q^*_{0,2}$. On the other hand, for any $\hat{\hat{Q}} \in \mathbb{Q}_{\theta,2}$, the first-order condition implies that $\varepsilon_\theta(\hat{\hat{Q}}) = 1$. As a result, it must be that $\hat{\hat{Q}} > Q^*_{0,2}$. Finally, if $\pi \sim$ Uni$[0,1]$, then since $v_\theta(\cdot)$ is strictly concave, the inverse demand curve $v_\theta \circ \hat{P}(Q)$ is also strictly concave. Since the iso-profit curve is strictly convex, there is at most one profit-maximizing quantity in $\mathbb{Q}_{\theta,2}$. □

LEMMA B.5. *Consider a non-degenerate SPNE of the upselling game, such that $P_1, P_2 \geq 0$ are the equilibrium prices for the basic product and the add-on.*

1. *If type $\bar{\pi}$ purchases both products, then any $\pi' > \bar{\pi}$ finds it strictly optimal to do so.*
2. *If type $\bar{\pi}$ only purchases the basic product in $t = 1$, then any $\pi' > \bar{\pi}$ at least purchases the basic product in the equilibrium.*

PROOF. For Part 1 of Lemma B.5, notice that type $\bar{\pi}$ purchases both the basic product and the add-on only if the following conditions are satisfied:

- The *ex-ante* self prefers purchasing both products over buying neither;
- After acquiring the basic product, the Period-2 self prefers purchasing the add-on rather than foregoing it.

If both conditions are met for the lower type $\bar{\pi}$, they will also be met for any higher type π'.

For Part 2 of Lemma B.5, suppose for a contradiction that $\bar{\pi}$ only purchases the basic product in $t = 1$, while for some type $\pi' > \bar{\pi}$, π' purchases neither of the products. First, notice that this scenario will *never* happen if $\theta = 0$, that is, if the consumers are dynamically consistent. The only reason for the higher type π' NOT to purchase the basic product, is that she foresees that her future self will purchase the add-on:

$$v_\theta(\pi') \geq P_2, \tag{53}$$

while purchasing both products is worse than purchasing nothing for her current self:

$$(1 + \alpha)\pi' < P_1 + P_2.$$

Let $\pi'' > \pi'$ be the lowest type who purchases both products in the equilibrium. Since the equilibrium is non-degenerate, it holds that $\pi'' \in (\pi', 1)$. For type π'', (53) and that $\pi'' > \pi'$

imply that

$$v_\theta(\pi'') > P_2. \tag{54}$$

Recall that the seller cannot commit to any period-2 price P_2 *ex-ante*. On the other hand, (54) implies that there is a profitable deviation for the period-2 seller to charge ϵ higher than P_2, such that (54) is still non-binding if P_2 is alternated by $P_2 + \epsilon$.

Given the consumers with type in the range $[\pi'', 1]$ have already purchased the basic product, from the period-2 seller's perspective, raising the add-on price by ϵ is profitable. These consumers will still purchase the add-on at the higher price, ensuring increased revenue without losing their business. Therefore, P_1 and P_2 cannot be equilibrium prices, a contradiction. □

LEMMA B.6. *For any non-degenerate SPNE of the upselling game, it holds that*

$$Q^*_{\theta,1} \geq \min \mathbb{Q}_{\theta,2}.$$

PROOF. Suppose for a contradiction that the period-1 sell $Q^*_{\theta,1} < \min \mathbb{Q}_{\theta,2}$. As in Definition B.1, $\mathbb{Q}_{\theta,2}$ is the collection of revenue-maximizing quanitities according to the inverse demand curve $v_\theta \circ \hat{P}(Q)$. As a result, $Q^*_{\theta,1} < \min \mathbb{Q}_{\theta,2}$ implies that $\varepsilon_\theta(Q^*_{\theta,1}) > 1$. By Lemma B.5, in the equilibrium, a consumer with type π purchase the basic product if and only if $\pi \in [0, Q^*_{\theta,1}]$. This is the consumer base for the seller at time $t = 2$ (recall that a consumer quits the market if she does not buy the basic product). For any $Q \in [0, Q^*_{\theta,1}]$, it holds that $\varepsilon_\theta(Q^*_{\theta,1}) > 1$; therefore, it is optimal for the period-2 seller to sell the add-on to everyone in $[0, Q^*_{\theta,1}]$, *i.e.*

$$Q^*_{\theta,2} = Q^*_{\theta,1} < \min \mathbb{Q}_{\theta,2}.$$

Correspondingly, at time $t = 2$, the seller sets the price at $P^*_{\theta,2} = v_\theta \circ \hat{P}(Q^*_{\theta,2}) > \alpha \hat{P}(Q^*_{\theta,2})$. The customers are forward-looking, therefore, $P^*_{\theta,1} + P^*_{\theta,2} \leq (1+\alpha)\hat{P}(Q^*_{\theta,2})$, which implies $P^*_{\theta,1} < \hat{P}(Q^*_{\theta,2})$; that is, the equilibrium price for the basic product is strictly lower than the *ex-ante* willingness-to-buy of the marginal customer at quantity level $Q^*_{\theta,2}$. However, this would imply $Q^*_{\theta,1} > Q^*_{\theta,2}$, which contradicts with $Q^*_{\theta,2} = Q^*_{\theta,1}$. Therefore, it must holds that $Q^*_{\theta,1} \geq \min \mathbb{Q}_{\theta,2}$. □

LEMMA B.7. *For any non-degenerate SPNE of the upselling game, it holds that* $Q^*_{0,1} = Q^*_{0,2}$*, moreover,*

$$Q^*_{\theta,2} > Q^*_{0,2},$$

and

$$Q^*_{\theta,1} > Q^*_{\theta,2}.$$

PROOF. First, $Q^*_{0,1} = Q^*_{0,2}$ since for each consumer, the valuation for the add-on is a fixed fraction α of the valuation of the basic product. Then notice that by Lemma B.6, $Q^*_{\theta,1} \geq \min \mathbb{Q}_{\theta,2}$. Since $\mathbb{Q}_{\theta,2}$ is defined to be the set of profit maximizing quantities, it is optimal for the period-2 seller to choose some $Q^*_{\theta,2} \in \mathbb{Q}_{\theta,2}$. By Lemma B.5, $Q \in \mathbb{Q}_{\theta,2}$ implies $Q > Q^*_{0,2}$. Therefore, $Q^*_{\theta,2} > Q^*_{0,2}$. Finally, $Q^*_{\theta,1} \geq Q^*_{\theta,2}$ mechanically holds, while assuming $Q^*_{\theta,1} = Q^*_{\theta,2}$ would lead to a contradiction, as illustrated in the proof of Lemma B.6. As a result, $Q^*_{\theta,1} > Q^*_{\theta,2}$. □

B.2.2 *Proof of Proposition 5.2* In this section, I assume that the type $\pi \sim \text{Uni}[0,1]$. In the analysis above, I characterized the properties of equilibrium *quantity* through analyzing the auxiliary inverse demand function $v_\theta \circ \hat{P}(Q)$ for the add-on. Now, to analyze the behavior of equilibrium *price*, I consider the corresponding demand function for the add-on, $D_\theta(P) = [v_\theta \circ \hat{P}(Q)]^{-1}$. Since $\pi \sim \text{Uni}[0,1]$, $\hat{P}(Q)$ is affine, which implies that $D_0(P)$ is also affine. Recall that v_θ is strictly concave; as a result, the inverse demand function $v_\theta \circ \hat{P}(Q)$, as well as the demand function $D_\theta(P)$, are both strictly concave. Therefore, there exists a increasing and strictly concave transformation $\sigma : [0,1] \to [0,1]$, such that

$$D_\theta(P) = \sigma \circ D_0(P);$$

moreover, $\sigma(0) = 0$ and $\sigma(1) = 1$. For any price level P, the price elasticity of demand is given by

$$\begin{aligned}
\varepsilon_\theta(P) &\equiv -\frac{D'_\theta(P) \cdot P}{D_\theta(P)} \\
&= -\frac{[\sigma \circ D_0(P)]' \cdot P}{\sigma \circ D_0(P)} \\
&= -\frac{\sigma'(D_0(P)) \cdot D'_0(P) \cdot P}{\sigma(D_0(P))} \\
&= -\frac{\sigma'(D_0(P)) \cdot D_0(P)}{\sigma(D_0(P))} \cdot \frac{D'_0(P) \cdot P}{D_0(P)}
\end{aligned}$$

$$< -\frac{D_0'(P) \cdot P}{D_0(P)} \equiv \varepsilon_0(P).$$

The last inequality holds as $\frac{\sigma'(Q) \cdot Q}{\sigma(Q)} < 1$, which holds true since $\sigma(0) = 0$, $\sigma(1) = 1$, while σ is strictly concave. In the SPNE of the upselling game, since $Q^*_{\theta,2} \in \mathbb{Q}_{\theta,2}$ is profit-maximizing for the demand function $D_\theta(P)$, the equilibrium price $P^*_{\theta,2}$ is also profit-maximizing for the demand function $D_\theta(P)$. Since $\varepsilon_\theta(P) < \varepsilon_0(P)$ for all $P \in (0, \alpha)$, it holds that $P^*_{\theta,2} > P^*_{0,2}$. For all customers, the valuation of the add-on is a constant fraction α of the basic product. Therefore, $P^*_{0,2} = \alpha P^*_{0,1}$. Finally, $P^*_{\theta,1} < P^*_{0,1}$ as otherwise, the price for both the basic product and the add-on (weakly) exceeds the rational benchmark, which is contradictory to the result in Proposition 5.1, stating that the equilibrium has a larger quantity of sales than the rational benchmark.

B.2.3 *Proof of Proposition 5.3* Recall that Proposition 5.1 states that $Q^*_{\theta,2} > Q^*_{0,2} = Q^*_{0,1}$. Since consumers are forward-looking, to guarantee that $Q^*_{\theta,2} > Q^*_{0,2}$, the total price $P^*_{\theta,1} + P^*_{\theta,2}$ in the upselling game must be strictly lower than the total price $P^*_{0,1} + P^*_{0,2}$ in the rational benchmark. Therefore, more consumers purchase both the basic product and the add-on than in the rational benchmark, while all consumers pay a lower total price. As a result, the total consumer surplus must be strictly higher.

B.2.4 *Proof of Proposition 5.4* In this proposition, I assume again $\pi \sim$ Uni$[0, 1]$, which implies $Q^*_{0,1} = Q^*_{0,2} = 1/2$. I define $\Delta Q \equiv Q^*_{\theta,2} - Q^*_{0,2} = Q^*_{\theta,2} - 1/2$ and $\Delta P \equiv P^*_{\theta,2} - \alpha(1 - Q^*_{\theta,2})$, where $\alpha(1 - Q^*_{\theta,2})$ is the *ex-ante* willingness-to-buy of the marginal customer who purchase the add-on in the equilibrium. Notice that $\Delta P > \alpha \Delta Q$; this is because the add-on price $P^*_{\theta,2} > P^*_{0,2} \equiv \alpha/2$. As a result, $\Delta P > \alpha/2 - \alpha(1 - Q^*_{\theta,2}) = \alpha/2 - \alpha[1 - (\Delta Q + 1/2)] = \alpha \Delta Q$.

At time $t = 1$, suppose the seller charges $\hat{P}_{\theta,1} \equiv 1/2 - \Delta Q - \Delta P$, and I want to show that in the subgame that follows, there will be $\hat{Q}_{\theta,1} = 1 - \hat{P}_{\theta,1}$ amount of consumers that purchase the basic product at time $t = 1$, moreover, the equilibrium price and quantity for add-on sales will be uniquely given by $P^*_{\theta,2}$ and $Q^*_{\theta,2}$. First, if $\hat{Q}_{\theta,1} = 1 - \hat{P}_{\theta,1}$, then by the definition of $\hat{P}_{\theta,1}$, $\hat{Q}_{\theta,1} = 1/2 + \Delta Q + \Delta P > 1/2 + \Delta Q = Q^*_{\theta,2}$. As a result, in period $t = 2$, it is optimal for the seller to charge the price $P^*_{\theta,2}$ that maximizes the profit from add-on sales, which corresponds to the quantity $Q^*_{\theta,2}$. Moreover, by Lemma B.4, the profit maximizing $P^*_{\theta,2}$ is unique.

It remains to verify that the quantity sold for $t = 1$ is $\hat{Q}_{\theta,1}$. First, consider the customer with type $1 - Q^*_{\theta,2}$, who would be marginally willing to consume the add-on in $t = 2$ after purchasing the basic product. *Ex-ante*, her willingness to buy is $(1+\alpha)(1-Q^*_{\theta,2})$, while the total price charged is $\hat{P}_{\theta,1}+P^*_{\theta,2}=1/2-\Delta P-\Delta Q+\Delta P+\alpha(1-Q^*_{\theta,2})=(1+\alpha)(1-Q^*_{\theta,2})$. As a result, customer with type $1-Q^*_{\theta,2}$ (and any higher type) is indeed willing to purchase both products in the equilibrium. Second, consider a customer with type in the interval $[1-\hat{Q}_{\theta,1}, 1-Q^*_{\theta,2})$; such a customer has the willingness-to-buy for the basic product exceeding $\hat{P}_{\theta,1}$, which is higher the price charged for the basic product. Therefore, all customers with types greater than $1-\hat{Q}_{\theta,1}$ (at least) purchase the basic product, which is consistent with the period-1 sales quantity being $\hat{Q}_{\theta,1}$.

For the seller, choosing price $\hat{P}_{\theta,1}\equiv 1/2-\Delta Q-\Delta P$ leads to basic product sells quantity $1-\hat{P}_{\theta,1}=1/2+\Delta Q+\Delta P$ and equilibrium add-on price and quantity $P^*_{\theta,2}=\Delta P+\alpha(1/2-\Delta Q)$ and $Q^*_{\theta,2}=1/2+\Delta Q$. Since the choice $\hat{P}_{\theta,1}$ is feasible for the seller initially, the equilibrium profit must exceed the profit corresponding to period-1 price $\hat{P}_{\theta,1}$, which equals

$$(1/2-\Delta Q-\Delta P)\cdot(1/2+\Delta Q+\Delta P)+[\Delta P+\alpha(1/2-\Delta Q)]\cdot(1/2+\Delta Q).$$

by algebraic calculation, the expression above equals

$$\frac{1}{4}(1+\alpha)+\Delta P\left(\frac{1}{2}-\Delta Q-\Delta P\right)-(1+\alpha)(\Delta Q)^2.$$

Since $\Delta P>\alpha\Delta Q$, the expression above is strictly greater than

$$\frac{1}{4}(1+\alpha)+\Delta Q\left(\frac{\alpha}{2}-(1+2\alpha)\Delta Q-\alpha\Delta P\right)$$

Fix α and take $\theta\downarrow 0$; then dynamic inconsistency becomes minimal, and $\Delta P,\Delta Q\to 0$. As a result, the term $\left(\frac{\alpha}{2}-(1+2\alpha)\Delta Q-\alpha\Delta P\right)$ is positive for all sufficiently small θ. For all such small θ, the seller's profit in the upselling game is strictly higher than

$$\frac{1}{4}(1+\alpha),$$

the benchmark profit level given $\theta=0$.

B.3 *Proof of Results in Section 5.2*

For signal realization s, dissonance factor θ and income Y, I let $c_2(s,\theta,Y)\in\{c^{Y,o},c^{Y,\nu}\}$ define the equilibrium period-2 choice in the CD game.

LEMMA B.8. *For each $\theta \geq 0$ and $Y \geq 0$, there exists $T(\theta, Y) > 0$ such that for signal s and Bayesian posterior $\mu_1^s = (\mu_{1\omega}^s)_{\omega=o,\nu}$,*

$$c_2(s,\theta,Y) = \begin{cases} o & \textit{if}\, \mu_{1o}^s/\mu_{1\nu}^s > T(\theta,Y); \\ \nu & \textit{if}\, \mu_{1o}^s/\mu_{1\nu}^s < T(\theta,Y). \end{cases}$$

Moreover, $T(\theta, Y)$ is strictly decreasing in θ and strictly increasing in Y.

PROOF. By Proposition 3.1, for signal s and Bayesian posterior μ_1^s, the DM chooses o at time $t = 2$ if

$$\mu_{1o}^s \phi_\theta \left[u(Y+1) + \delta u(Y+1)\right] + \mu_{1\nu}^s \phi_\theta \left[u(Y) + \delta u(Y)\right]$$
$$> \mu_{1o}^s \phi_\theta \left[u(Y+1) + \delta u(Y)\right] + \mu_{1\nu}^s \phi_\theta \left[u(Y) + \delta u(Y+1)\right].$$

That is,

$$\mu_{1o}^s \phi_\theta[u(Y+1)] > \mu_{1\nu}^s \phi_\theta[u(Y)],$$

i.e.

$$\frac{\mu_{1o}^s}{\mu_{1\nu}^s} > \phi_\theta[-u(Y+1) + u(Y)].$$

Similarly, the DM chooses ν at time $t = 2$ if

$$\frac{\mu_{1o}^s}{\mu_{1\nu}^s} < \phi_\theta[-u(Y+1) + u(Y)].$$

Therefore, the $T(\theta, Y)$ in Lemma B.8 exists and $T(\theta, Y) = \phi_\theta[u(Y+1) - u(Y)]$. Since u is strictly concave, $T(\theta, Y)$ is decreasing in Y. Since $\phi_\theta(\cdot) = \exp[\theta(\cdot)]$, $T(\theta, Y)$ is increasing in θ. □

Notice that the probability of switching to the new technology at time $t = 2$ is equal to

$$Prob(\{s \mid \mu_{1o}^s/\mu_{1\nu}^s < T(\theta,Y)\}),$$

which is strictly increasing in $T(\theta, Y)$. As a result, the probability of switching to the new technology is strictly decreasing in θ and strictly increasing in Y.

B.4 *Proof of Results in Section 5.3*

In this section, I consider $X=\mathbb{R}$ and a fixed period-1 action $a=(c_1,C_2)$, where $C_2=\{c_2^i\}_{i=1}^n$ is finite and $V_a(\rho_0,\theta_0)\neq V_a(\rho_0,0)$ for some $\theta_0>0$ and $\rho_0\in\mathcal{B}$. Notice that may I transform action a to $\hat{a}=(0,\hat{C}_2)$ without influencing the value of information, where $0\in\mathbb{R}^\Omega$ and $\hat{C}_2=\{\frac{c_1}{\delta}+c_2\mid c_2\in C_2\}\subset\mathbb{R}^\Omega$. Formally, it holds that, for all $\rho\in\mathcal{B}$ and $\theta\geq 0$, $V_a(\rho,\theta)=V_{\hat{a}}(\rho,\theta)$. Therefore, without loss of generality, for the period-1 action $a=(c_1,C_2)$ fixed above, I assume $c_1=0\in\mathbb{R}^\Omega$.

For each $i=1,2,...,n$, $\theta\geq 0$, I define

$$P_{\theta,i}=\{\mu\in\Delta\Omega\mid(c_1,c_2^i)\in SPNE_\kappa(a),\ \text{where}\ \kappa=(\mu,u,\delta,\theta)\}.$$

as the collection of *ex-ante* beliefs such that c_2^i is an equilibrium period-2 choice for the CD game with dissonance factor θ. Let $\hat{I}_\theta=\{i=1,2,...,n\mid Int(P_{\theta,i})\neq\emptyset\}$ denote the collection of all $i=1,2,...,n$ such that c_2^i is chosen at $t=2$ for a positive "mass" of *ex-ante* beliefs.

LEMMA B.9. *If $\theta>\theta'$, then $\hat{I}_\theta\subset\hat{I}_{\theta'}$.*

PROOF. Take $c_2^0\in C_2$ such that $c_2^0\notin\hat{I}_{\theta'}$ and I want to show that $c_2^0\notin\hat{I}_\theta$. For $i=0,1,...,n$ and $c_2^i\in C_2$, define $v^i=(v^i_\omega)_{\omega\in\Omega}=\phi_{\delta\theta'}\circ u(c_2^i)$. Since $c_2^0\notin\hat{I}_{\theta'}$, it holds for all $\mu\in\Delta\Omega$ that

$$\mu\cdot v^0\leq\max_{i\in\hat{I}_{\theta'}}\mu\cdot v^i,$$

which, by the supporting hyperplane theorem, implies that $v^0\leq\sum_{i\in\hat{I}_{\theta'}}\lambda^i v^i$ for some $\lambda\in\Delta(\hat{I}_{\theta'})$. Now, define z^i such that for every $\omega\in\Omega$, $z^i_\omega=\phi_{\delta(\theta-\theta')}(v^i_\omega)$, then by Jensen's inequality, it holds that $z^0<\sum_{i\in\hat{I}_{\theta'}}\lambda^i z^i$, which implies that

$$\mu\cdot z^0<\max_{i\in\hat{I}_{\theta'}}\mu\cdot z^i.$$

Since this inequality holds for all μ, it holds that $c_2^0\notin\hat{I}_\theta$. □

Lemma B.9 establishes that the set $\hat{I}_\theta$ "shrinks" as θ increases. Thus, any item $c_2^i\in C_2$ with $i\notin\hat{I}_0$ for $\theta=0$ will *never* be chosen for any *ex-ante* belief for any dissonance factor $\theta>0$. Without loss of generality, we delete such irrelevant items and assume that, when $\theta=0$, it holds that $\hat{I}_0=\{1,2,\ldots,n\}$. That is, in the rational benchmark, any item in C_2 can be chosen for a positive "mass" of *ex-ante* beliefs.

Next, I introduce a definition that will be used extensively in the proof below.

DEFINITION B.2. *For $\theta \geq 0$, I define the value of belief $\mu \in \Delta\Omega$ by*

$$W_a(\mu,\theta) = \sup_{c \in SPNE_{\hat{\kappa}}(\{a\})} U_\mu(c).$$

By the definition, $V_a(\rho,\theta) = \mathbb{E}_\rho[W_a(\mu,\theta)] - W_a(\mu(\rho),\theta)$. As in the standard result, given that $\theta = 0$, $W_a(\mu,0)$ is weakly convex in μ over $\Delta\Omega$, which implies that the value of information is always weakly positive. However, in general, $W_a(\mu,\theta)$ is not convex in μ for $\theta > 0$. Also, notice that

$$W_a(\mu,\theta) = \sup_{i \in \{1,2,\dots,n\}: \mu \in P_{\theta,i}} \mathbb{E}_\mu[u(c_2^i)].$$

Moreover, if $\mu \in Int(P_{\theta,i})$ for some $i = 1,2,\dots,n$, then

$$W_a(\mu,\theta) = \mathbb{E}_\mu[u(c_2^i)].$$

B.4.1 *Proof of Proposition 5.6* Let $\bar{\theta} > 0$ be a dissonance factor sufficiently small that $Int(P_{0,i}) \cap Int(P_{\theta,i}) \neq \emptyset$ for all $i = 1,2,\dots,n$. Recall that I assume $V_a(\rho_0,\theta_0) \neq V_a(\rho_0,0)$ for some $\rho_0 \in \mathcal{B}$ and $\theta_0 > 0$. As a result, for all $\theta \in [0,\bar{\theta})$, there exists $j \in \{1,2,\dots,n\}$ such that $Int(P_{0,j}) \neq Int(P_{\theta,j})$. Take $\hat{\rho} \in \mathcal{B}$ such that $Supp(\rho) \subset Int(P_{0,i})$, $\mu(\rho) \in Int(P_{0,i}) \cap Int(P_{\theta,i})$ and $Prob(\hat{\rho} \notin P_{0,i}) > 0$.

- Since $Supp(\hat{\rho}) \subset Int(P_{0,i})$, it holds that $\mathbb{E}_{\hat{\rho}}[W_a(\mu,0)] = W_a(\mu(\hat{\rho}),0)$.
- Since $\mu(\hat{\rho}) \in Int(P_{0,i}) \cap Int(P_{\theta,i})$, it holds that $W_a(\mu(\hat{\rho}),\theta) = W_a(\mu(\hat{\rho}),0)$.
- Since $Prob(\hat{\rho} \notin P_{0,i}) > 0$, it holds that $\mathbb{E}_{\hat{\rho}}[W_a(\mu,\theta)] < \mathbb{E}_{\hat{\rho}}[W_a(\mu,0)]$.

Therefore, $V_a(\hat{\rho},0) = 0$ while $V_a(\hat{\rho},\theta) < 0$. It holds that

$$V_a(\hat{\rho},\theta) < 0 \leq V_a(\hat{\rho},0).$$

Now, take $\hat{\rho}' \in \mathcal{B}$ such that $\mu(\hat{\rho}') \in Int(P_{0,j})$ but $\mu(\hat{\rho}') \notin P_{\theta,j}$ and $Supp(\hat{\rho}') \in \{\delta_\omega\}_{\omega \in \Omega}$ where δ_ω is the Dirac delta for $\omega \in \Omega$.

- Since $\mu(\hat{\rho}') \in Int(P_{0,j})$ but $\mu(\hat{\rho}') \notin P_{\theta,j}$, it holds that $W_a(\mu(\hat{\rho}'),0) > W_a(\mu(\hat{\rho}'),\theta)$.

- Notice that the CD game does not create belief distortion if *ex-ante* belief is degenerate. Since $Supp(\hat{\rho}') \in \{\delta_\omega\}_{\omega\in\Omega}$, it holds that it holds that $\mathbb{E}_{\hat{\rho}'}[W_a(\mu,0)] = \mathbb{E}_{\hat{\rho}'}[W_a(\mu,\theta)]$.

Therefore, $V_a(\hat{\rho}',\theta) > V_a(\hat{\rho}',0) > 0$.

For the remaining propositions, I consider $\Omega = \{L,R\}$. Without loss of generality, assume $C_2 = \{c_2^i\}_{i=1}^n$, where c_{2L}^i is strictly increasing in i and c_{2R}^i is strictly decreasing in i.[42]

B.4.2 *Proof of Proposition 5.7* Take θ^* sufficiently large that $P_{\theta,1} \cup P_{\theta,n} = [0,1]$ for all $\theta \geq \theta^*$.

- Case 1 $c_{2R}^1 = c_{2L}^n$ and $c_{2L}^1 = c_{2R}^n$. Similar to the proof of Proposition C.2 above, $V_a(\rho,\theta) \geq 0$ for all $\rho \in \mathcal{B}$ and $\theta > \theta^*$. Therefore, $P_{Neg}(a,\theta) = \emptyset$ for all $\theta \geq \theta^*$.

- Case 2 $c_{2R}^1 = c_{2L}^n > c_{2L}^1 > c_{2R}^n$. In this case, for any $\theta \geq \theta^*$, it can be shown that $\sup P_{\theta,1} > \frac{1}{2}$ and $\sup P_{\theta,1}$ decreases with θ. Moreover, it can be shown that $P_{Neg}(a,\theta) = [0, \sup P_{\theta,1})$. As a result, $P_{Neg}(a,\theta)$ shrinks with increasing θ.

- Case 3 $c_{2R}^1 > c_{2L}^n > \max\{c_{2L}^1, c_{2R}^n\}$. In this case, for θ sufficiently large, it can be shown that $\sup P_{\theta,1}$ increases with θ and converges to 1. Moreover, it can be shown that $P_{Neg}(a,\theta) = (\sup P_{\theta,1}, 1]$. Therefore, $P_{Neg}(a,\theta)$ shrinks with increasing θ.

[42] That is, any option that are weakly statewisely dominated by some other option is removed from C_2.

Appendix C: Miscellaneous

C.1 *Past Actions and Undesirable Consequences*

Researches	Past Actions	Undesirable Outcomes
Akerlof and Dickens (1982)	Choosing a risky job	Workplace accident
Rabin (1994), Konow (2000)	A morally dubious action	Punishment or guilt
Oxoby (2004)	Exerting work effort	Low income and low status spending
Mullainathan and Washington (2009)	Voting for a politician	The politician is low quality
Chang et al. (2016)	Investing in a risky asset	The investment is unprofitable
Penn (2017)	Acquiring a certain skill	The skill is useless
Acharya et al. (2018)	Supporting a political party	The party advocates flawed policies
Fan (2024)	Improving a certain attribute of an investment	The attribute is irrelevant to investment outcome

Table C.1. Examples of past actions and undesirable consequences

C.2 *Cognitive Tax on Poverty – Excessive Risk-taking*

Consider a setting that slightly modifies Section 5.2. Same as Section 5.2, the DM starts with using the old technology o, receives a signal realization from the information structure, and distorts the Bayesian posterior. Different from the setting above, instead of $\{o, \nu\}$, the period-2 menu is given by $\{o, O\}$. The state-dependent payoff of each choice in $\{o, O\}$ is represented by a pair (c_o, c_ν), with $o = (Y+1, Y)$ and $O = (Y+1+\epsilon, Y-\epsilon)$. Here, O represents making a greater investment in the old technology at time $t = 2$. For $c_2 \in \{o, O\}$, I let $P^*_\epsilon(c_2 \mid \theta, Y)$ define the probability that option c_2 is chosen at time $t = 2$.

Proposition C.1. *(Escalation of Commitment) For $\theta > 0$ and $Y < Y'$, there exists a sufficiently small ϵ^* such that for all $\epsilon < \epsilon^*$,*

$$P^*_\epsilon(O \mid \theta, Y) > P^*_\epsilon(O \mid \theta, Y').$$

Proposition C.1 states that when the additional investment ϵ is sufficiently small, an individual with a lower baseline income Y is more likely to escalate their prior commitment to the old technology. Specifically, it is more likely that she chooses the investment option O, which involves taking greater risks through making a larger investment in the old technology.

C.3 *Cognitive Dissonance and Information Avoidance*

For $\Omega = \{L, R\}$, $a = (c_1, C_2)$ is *state-neutral* if $c_1 = (0,0)$ and $c_2 = (c_{2L}, c_{2R}) \in C_2 \implies c_2' = (c_{2R}, c_{2L}) \in C_2$. This corresponds to a situation where the DM perceives no inherent difference between L and R.

PROPOSITION C.2. *Suppose* $|\Omega| = 2$, a *is regular and state-neutral, then for every* $\theta > 0$,

- *the closure of* $P_{Neg}(a,\theta)$ *is either empty or equals* $[\epsilon_\theta, 1-\epsilon_\theta]$ *for some* $\epsilon_\theta \in \left(0, \frac{1}{2}\right]$;
- *the Lebesgue measure of* $P_{Neg}(a,\theta)$ *decreases with* θ *over* $(0,\infty)$ *with* $P_{Neg}(a,\theta) = \emptyset$ *for all sufficiently large* θ.

Proposition C.2 formalizes the intuition discussed in Section 5.3. It states that information may only hurt people with *weaker* priors (those farther away from certainty), who tend to choose less risky options with no information available. Moreover, $P_{Neg}(a,\theta)$ is empty for all sufficiently large θ, meaning information never hurts when the desire for dissonance mitigation is sufficiently strong.

PROOF. For $\theta \geq 0$ and $i = 1, 2, ..., n$, the set of beliefs $P_{\theta,i}$ is the collection of *ex-ante* beliefs such that c_2^i is an equilibrium period-2 choice for the CD game. For $\theta \geq 0$, take $\epsilon_\theta \in (0, \frac{1}{2}]$ such that $P_{\theta,1} = [0, \epsilon_\theta]$ and $P_{\theta,n} = [1-\epsilon_\theta, 1]$. The size of $P_{\theta,1}$ and $P_{\theta,n}$ are identical since a is state-neutral.

Take $\theta > 0$ such that $P_{Neg}(a,\theta) \neq \emptyset$. I want to show that the closure of $P_{Neg}(a,\theta)$ equals $[\epsilon_\theta, 1-\epsilon_\theta]$. Take any $\hat{i} \in \{2, 3, ..., n-1\}$ and $\hat{\mu} \in Int(P_{\theta,i})$.[43] It suffices to show that there exists an experiment $\hat{\rho}$ such that $\mu(\hat{\rho}) = \hat{\mu}$ and $V_a(\hat{\rho}, \theta) < 0$.

Without loss of generality, assume that $\hat{i} \leq \frac{1}{2}(n+1)$, which implies that $P_{\theta,\hat{i}} \subset [0, \frac{1}{2}]$. In this case, it must hold that $\inf P_{\theta,\hat{i}} > \inf P_{0,\hat{i}}$, which implies that $\lim_{\mu \downarrow \inf P_{\theta,\hat{i}}} W_a(\mu,\theta) - \lim_{\mu \uparrow \inf P_{\theta,\hat{i}}} \equiv G > 0$. Now, construct experiment $\{\rho^n\}_{n=1}^\infty$ with $Supp(\rho^n) = \{\inf P_{\theta,\hat{i}} - 1/n, \sup P_{\theta,\hat{i}}\}$. Take $n \to \infty$, it holds that $\lim_{n\to\infty} \mathbb{E}_{\rho^n} W_a(\mu,\theta) = W_a(\hat{\mu},\theta) - \frac{\sup P_{\theta,\hat{i}} - \hat{\mu}}{\sup P_{\theta,\hat{i}} - \inf P_{\theta,\hat{i}}} G$, i.e. $\lim_{n\to\infty} V_a(\rho^n,\theta) = -\frac{\sup P_{\theta,\hat{i}} - \hat{\mu}}{\sup P_{\theta,\hat{i}} - \inf P_{\theta,\hat{i}}} G < 0$. As a result, for n sufficiently large, it holds that $V_a(\rho^n,\theta) < 0$.

[43] Such $\hat{i}$ must exist; otherwise, $P_{Neg}(a,\theta) = \emptyset$.

Finally, for $i \in \left(1, \frac{n+1}{2}\right]$, define $Q_{\theta,i} = \{\mu \mid \mathbb{E}_\mu[\phi_{\delta\theta} \circ u(c_1)] \geq \mathbb{E}_\mu[\phi_{\delta\theta} \circ u(c_i)]\}$. By definition, $P_{\theta,1} = \bigcap_{i \in \left(1, \frac{n+1}{2}\right]} Q_{\theta,i}$. Observe that for each i, the set $Q_{\theta,i}$ expands as θ increases. Consequently, $P_{\theta,1}$ also expands with increasing θ, meaning ϵ_θ grows with θ. Furthermore, for each i, $Q_{\theta,i} = [0, \frac{1}{2}]$ when θ is sufficiently large. Hence, $P_{\theta,1} = [0, \frac{1}{2}]$ for sufficiently large θ. By symmetry, $P_{\theta,n} = \left[\frac{1}{2}, 1\right]$ for sufficiently large θ. In this case, $V_a(\rho,\theta) \geq 0$ for all $\rho \in \mathcal{B}$, which implies $P_{Neg}(a,\theta) = \emptyset$. $\square$

PROPOSITION C.3. *Suppose* $|\Omega| = 2$. *For every regular* a *and* $\theta > 0$, *there exists* $0 < \underline{\epsilon}_\theta < \bar{\epsilon}_\theta < 1$, *such that for* $\rho \in \mathcal{B}$,

$$Supp(\rho) \in [0, \underline{\epsilon}_\theta] \cup [\bar{\epsilon}_\theta, 1] \implies V_a(\rho,\theta) \geq V_a(\rho,0).$$

Moreover, for some ρ' *such that* $Supp(\rho') \in [0, \underline{\epsilon}_\theta] \cup [\bar{\epsilon}_\theta, 1]$, *it holds that* $V_a(\rho',\theta) > V_a(\rho',0)$.

PROOF. Take $\underline{\epsilon}_\theta = \sup[P_{0,1} \cap P_{\theta,1}]$ and $1 - \bar{\epsilon}_\theta = \inf[P_{0,n} \cap P_{\theta,n}]$. By construction, for any $\mu \in [0, \underline{\epsilon}_\theta] \cup [1 - \bar{\epsilon}_\theta, 1]$, it holds that $W_a(\mu,\theta) = W_a(\mu,0)$. Therefore, for any $\rho \in \mathcal{B}$ such that $Supp(\rho) \in [0, \underline{\epsilon}_\theta] \cup [1 - \bar{\epsilon}_\theta, 1]$, it holds that $\mathbb{E}_\rho W_a(\mu,\theta) = \mathbb{E}_\rho W_a(\mu,0)$. Moreover, for $\mu(\rho)$, it must hold that $W_a(\mu,\theta) \leq_a (\mu,0)$ since $W_a(\mu,\theta)$ is weakly decreasing in θ. Therefore, $V_a(\rho,\theta) \geq V_a(\rho,0)$. Since a is regular, there exists μ^* such that $W_a(\mu^*,\theta) < W_a(\mu^*,0)$. Take experiment ρ' with $\mu(\rho') = \mu^*$ and $Supp(\rho') \in [0, \underline{\epsilon}_\theta] \cup [1 - \bar{\epsilon}_\theta, 1]$, then it must hold that $V_a(\rho,\theta) > V_a(\rho,0)$. $\square$