Three Essays in Decision Theory

by

Elchin Suleymanov

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Economics)
in The University of Michigan
2019

Doctoral Committee:

Professor Yusufcan Masatlioglu, University of Maryland, Co-Chair
Assistant Professor Shaowei Ke, Co-Chair
Professor Matias D. Cattaneo
Associate Professor David A. Miller
Professor Rocio Titiunik

Elchin Suleymanov

elchin@umich.edu

ORCID iD: 0000-0003-4816-5886

ACKNOWLEDGMENTS

I am indebted to Yusufcan Masatlioglu for his constant support over the past years. Without his advice and encouragement, I would not be able to complete my dissertation. I thank Shaowei Ke for his support and guidance, especially during my last year at the University of Michigan. The first chapter of this dissertation would not exist without his encouragement and patient advice. I thank Matias Cattaneo, David Miller, and Rocio Titiunik for their advice and feedback. I have benefited from discussions with numerous other people during my undergraduate and graduate studies. I apologize that the list is too long for me to be able to mention everyone. I thank my officemates Xing Guo, Xinwei Ma, and Kenichi Nagasawa for helpful discussions and a productive work environment. Finally, I thank my parents Aliye Suleymanova and Tofiq Suleymanov for their support and care.

TABLE OF CONTENTS

LIST OF FIGURES

FIGURE

ABSTRACT

This dissertation contributes to a few topics in decision theory including non-Bayesian updating, reference dependent choice, and limited attention. In the first chapter, I propose and axiomatically characterize a novel belief updating rule, called *robust maximum likelihood updating*, which can accommodate most commonly observed biases in probabilistic reasoning. In this updating rule, the decision maker is endowed with a *benchmark prior*, which is interpreted as the decision maker's "best guess," and a *set of plausible priors*. When the decision maker receives new information, she first revises her benchmark prior and then performs Bayesian updating based on the new benchmark prior. The revision of the benchmark prior is done in two stages. First, the decision maker restricts her attention to the subset of plausible priors which maximize the likelihood of the observed information. Next, the decision maker chooses a new prior within this set that is as "close" to the original benchmark prior as possible. The first stage reflects the decision maker's willingness to learn from the new information, and the second stage reflects her willingness to be as *dynamically consistent* as possible. I take the decision maker's preferences over acts (e.g., bets) before and after the arrival of new information as the primitive of analysis and propose axioms which characterize this updating rule. I show that the model provides explanations for confirmation bias, base rate neglect, conservatism bias and overconfidence.

In the second chapter, I investigate choice behavior that differs from the standard rational choice model in two ways: (i) the decision maker's choices are reference dependent, and (ii) the decision maker pays attention to a *subset* of available alternatives. In this framework, I provide novel axiomatic characterizations of three commonly used random attention models: fixed independent consideration, logit consideration, fixed correlated consideration. While these models, or their

variations, have been previously examined in the literature, the relationship between these models was not precisely known. My axiomatic characterization makes this relationship clear. First, I show that the fixed independent consideration model can be characterized by two key properties: *irrelevance of dominated alternatives* and *ratio independence of dominant alternatives*. Next, I show that logit consideration relaxes the former property while fixed correlated consideration relaxes the latter. Hence, the intersection of logit and fixed correlated consideration models is exactly fixed independent consideration. Finally, I illustrate how attention parameters can be (partially) recovered from observed choice behavior in all these models.

In the third chapter which is coauthored with Yusufcan Masatlioglu, we study a model of search within a *product network*. A product network consists of a vast number of goods which are linked to one another. We investigate decision making in this new environment by using revealed preference techniques. In our model, the decision maker searches within the product network to uncover available goods. Due to the constraint imposed by the network structure and the starting point of search, the decision maker might not discover all available goods. We illustrate how one can deduce both the decision maker's preference and her product network from observed behavior. We also consider an extension of the model where the decision maker terminates the search before exhausting all the options (limited search).

CHAPTER 1

Robust Maximum Likelihood Updating

1.1 Introduction

How do decision makers (DMs) update their beliefs when they receive new information? The answer to this question is critical in economic models and policy analyses where one tries to predict the consequence of releasing new information to market participants. The standard assumption in economics is that beliefs are updated using Bayes' rule. However, there is a large body of experimental and empirical evidence which shows that decision makers frequently deviate from Bayesian updating. For example, many decision makers tend to underweight base rates (*base rate neglect*), ignore informative signals (*conservatism*), interpret contrary evidence as supportive of their original beliefs (*confirmation bias*).[1] This paper introduces a model where deviations from Bayesian updating are due to ambiguity/multiple priors. The model can accommodate previously mentioned and other errors in probabilistic reasoning.

To illustrate how deviations from Bayesian updating can be related to multiple priors, consider the thought experiment due to Ellsberg (1961) where a DM is told that an urn contains 30 red balls and 60 blue or green balls in an unknown proportion. Let f_R, f_B, and f_G stand for bets which yield \$100 if the ball drawn from the urn is red, blue, and green, respectively, and \$0 otherwise. When no further information is given, many decision makers are indifferent between these bets, which is consistent with the prior that assigns equal probability to all colors.[2]

[1]For a review of these findings, see, for example, Camerer (1995), Rabin (1998), Tversky (2004), and Benjamin (2019).

[2]Ellsberg argued that most decision makers would prefer to bet on red rather than blue or green. While Ellsberg style preferences are common, many experimental findings show that a significant number of decision makers are

Now suppose the experimenter draws a ball from the urn and conveys to the DM that the ball is not green. How should the DM update her preferences given this information? In particular, should she still be indifferent between f_R and f_B? There are two arguments that can be made. First, following the principle of *dynamic consistency*, one can argue that since both f_R and f_B agree on the payoff assigned to the unrealized event (green), the information that this event is ruled out should not affect the original preference. Hence, indifference should be maintained ex post. On the other hand, the information that the ball is not green may suggest that the number of blue balls in the urn is greater than the number of green balls. Since there are only 30 red balls in the urn and 60 blue or green balls, one can also argue that f_B should be preferred to f_R ex post. This preference is incompatible with dynamic consistency, which is the key implication of Bayesian updating.

The intuition that decision makers may not perform Bayesian updating when they face ambiguity is confirmed by experiments and observations of practitioners' behavior. For example, in a similar dynamic Ellsberg experiment, Dominiak, Duersch, and Lefort (2012) find that a significant number of decision makers whose behavior can be characterized as ambiguity neutral are not Bayesian. In addition, many statistical tools used in practice (e.g., maximum likelihood estimation, hypothesis testing, etc.) are non-Bayesian even though conceivably many statisticians, and scientists in general, may be ambiguity neutral.

Most existing models tie the DM's ambiguity attitude (rather than ambiguity) to her response to new information, which forces an ambiguity neutral DM to update her beliefs using Bayes' rule. This is not only inconsistent with the intuition and observations described above but also unnatural as ambiguity attitude and belief updating are distinct concepts. The model proposed in this paper allows the DM depart from Bayesian updating when she faces ambiguity even if her attitude towards ambiguity is neutral.

I adopt a dynamic version of the classical Anscombe and Aumann (1963) setup. Let Ω be a finite set of states, and denote by $\Delta(\Omega)$ the set of all probability measures on Ω. An event is

ambiguity neutral (see Binmore, Stewart, and Voorhoeve, 2012; Charness, Karni, and Levin, 2013; Stahl, 2014). In this paper, I consider both ambiguity neutral and ambiguity averse decision makers.

a member of $\mathcal{A}$, which is the collection of all subsets of Ω. The set of prizes (e.g., monetary payments) is a convex subset of a metric linear space, and an act is a function that assigns a prize to each state of the world.

The primitive of the analysis is a collection of preferences $\{\succsim_A\}_{A\in\mathcal{A}}$ where $\succsim_A$ represents the DM's preference over acts when she learns that event A occurs. The preference when the DM receives no information is $\succsim_\Omega$, which, for simplicity, is denoted by $\succsim$. The main axioms in this paper characterize a representation where the DM is endowed with *a benchmark prior* $\pi \in \Delta(\Omega)$, revealed from ex ante preferences $\succsim$, and *a set of plausible priors* $\mathbb{N}(\pi)$, revealed from updated preferences $\succsim_A$. For example, in the Ellsberg experiment the benchmark prior may assign equal probability to all colors, while any prior that assigns 1/3 probability to red is plausible. The benchmark prior π is interpreted as the DM's initial best guess where $\pi \in \mathbb{N}(\pi)$.

The axioms yield a novel updating rule, *robust maximum likelihood (RML) updating*, which can be described by two stages. In the first stage, the DM performs maximum likelihood updating within the set of plausible priors. That is, when the DM learns that an event A occurs, she restricts her attention to the subset of plausible priors which maximize the likelihood of this event. This set is denoted by

$$\mathbb{N}_A(\pi) = \arg\max_{\pi' \in \mathbb{N}(\pi)} \pi'(A).$$

Next, the DM chooses a new benchmark prior that induces maximally dynamically consistent behavior among all priors in $\mathbb{N}_A(\pi)$ and updates it using Bayes' rule. Maximal dynamic consistency ensures that the DM stays as "close" to her original benchmark prior as possible. A similar idea is also used in robust control literature where the benchmark prior π is treated as an "approximating model" that is not fully trusted, and models "further away" from π are seen as less appealing (see Hansen and Sargent, 2001; Strzalecki, 2011). When there is no ambiguity (i.e., $\mathbb{N}(\pi)$ is a singleton), RML updating reduces to Bayesian updating.

To illustrate how RML updating can accommodate a strict preference for f_B over f_R after the realization that the ball drawn from the Ellsberg urn is not green, let $\pi, \pi', \pi'' \in \mathbb{N}(\pi)$ where π is the benchmark prior, and π' and π'' represent two plausible priors when there are no green and blue balls in the urn, respectively.

	Red	Blue	Green
π	$1/3$	$1/3$	$1/3$
π'	$1/3$	$2/3$	0
π''	$1/3$	0	$2/3$

When the DM learns that the ball drawn from the urn is not green, maximum likelihood updating implies $\mathbb{N}_A(\pi) = \{\pi'\}$. The DM endowed with π' as her posterior prefers f_B over f_R.

RML updating provides explanations for most commonly observed biases in probabilistic reasoning. For example, consider confirmation bias, which is the tendency to interpret contrary evidence as supportive of original beliefs (see Rabin and Schrag, 1999, and references therein). Let $S = \{s_1, s_2\}$ be the set of payoff-relevant states, and denote by $\Sigma = \{\sigma_1, \sigma_2\}$ the set of signals. Suppose the DM assesses s_1 to be more likely than s_2, and σ_i is considered more likely than σ_j when the payoff-relevant state is s_i. The joint state space Ω and the benchmark prior π are illustrated below where $\mu > 1/2$ and $\alpha > 1/2$. A decision maker who displays confirmation bias might believe that s_1 is at least as likely as before when she observes σ_2.

	σ_1	σ_2
s_1	ω_{11}	ω_{12}
s_2	ω_{21}	ω_{22}

State Space Ω

	σ_1	σ_2
s_1	$\mu\alpha$	$\mu(1-\alpha)$
s_2	$(1-\mu)(1-\alpha)$	$(1-\mu)\alpha$

Benchmark Prior π

Imagine that the decision maker is not fully confident in the link between s_1 and the signals given by her benchmark prior and finds it plausible that the true information structure is given by π' (illustrated below) where $\alpha' < \alpha$. For example, the DM might find it plausible that there is a "bias" in the information source that is potentially unfavorable towards the state she originally finds more likely (i.e., the state s_1). Now suppose the DM observes the realization σ_2. Notice that under the benchmark prior the probability of observing σ_2 is $\mu + \alpha - 2\mu\alpha$. On the other hand, under the alternative plausible prior the probability of observing σ_2 is $\mu + \alpha - \mu\alpha - \mu\alpha'$. Hence, after observing σ_2 the DM performing maximum likelihood updating may change her benchmark prior from π to π'. This will result in a behavior that is consistent with confirmation bias.

	σ_1	σ_2
s_1	$\mu\alpha'$	$\mu(1-\alpha')$
s_2	$(1-\mu)(1-\alpha)$	$(1-\mu)\alpha$

Alternative Plausible Prior π'

An important question is whether one can identify the benchmark prior π and the set of plausible priors $\mathbb{N}(\pi)$ from preferences. The identification of π from ex ante preferences can be done as in Anscombe and Aumann (1963). To illustrate how $\mathbb{N}(\pi)$ can be identified, I first distinguish between *unambiguous* and *ambiguous* events. An event is unambiguous if there is full agreement among all plausible priors on its likelihood.[3] Otherwise, it is ambiguous. A prior is considered plausible if and only if it agrees with the benchmark prior on the likelihood of all unambiguous events.

Since the main axioms in this paper imply subjective expected utility (SEU) preferences, most existing approaches in the literature do not help us identify unambiguous events from preferences.[4] A novelty in this paper is that unambiguous events are identified by comparing ex ante and ex post preferences. To see how this can be done, notice that the DM may not satisfy dynamic consistency when an ambiguous event is realized, as illustrated by the Ellsberg example. Dynamic consistency requires that if two acts f, g agree outside an event E and f is ex ante preferred to g, then f must still be preferred to g when E is realized. Formally, $f(\omega) = g(\omega)$ for all $\omega \in E^c$ and $f \succsim g$ imply $f \succsim_E g$. Since all plausible priors agree on the likelihood of unambiguous events, preferences are expected to satisfy dynamic consistency when an unambiguous event occurs. More importantly, consider an event $B \supseteq E$ where E is unambiguous. Since E is unambiguous and f, g agree on

[3] Several papers have provided a behavioral definition for unambiguous events. Epstein and Zhang (2001), Zhang (2002), and Gul and Pesendorfer (2014) define unambiguous events to be the ones which satisfy some versions of Savage's Sure Thing Principle. Ghirardato, Maccheroni, and Marinacci (2004) argue for a "relation based" approach and provide a definition for an act to be unambiguously preferred to another act.

[4] To be more precise, most existing approaches use only ex ante preferences to reveal ambiguity (see Epstein, 1999; Ghirardato and Marinacci, 2002). Hence, using the standard terminology, the DM whose ex ante preferences satisfy SEU axioms can be characterized as ambiguity neutral. The main difference in this paper is that an ambiguity neutral DM may still have multiple priors which will be reflected in ex post preferences even though it is not reflected in ex ante preferences.

$B \backslash E$, one also expects that the DM's ex ante preference between f and g should be preserved when B is realized. According to this observation, E is defined to be *perfectly dynamically consistent* if for any two acts f, g that agree on E^c and any event $B \supseteq E$, f is ex ante preferred to g if and only if f is preferred to g when B is realized.

To identify $\mathbb{N}(\pi)$ from preferences, I first define an event E to be unambiguous when both E and E^c are perfectly dynamically consistent. The set $\mathbb{N}(\pi)$ consists of all probability measures on Ω which agree with the benchmark prior π on the likelihood of all unambiguous events. In the Ellsberg example, if the DM is indifferent between f_R and f_B ex ante but strictly prefers f_B to f_R when she is told that the ball drawn from the urn is not green, the definition implies that both $\{R, B\}$ and $\{G\}$ are ambiguous events, and hence there are multiple plausible priors which differ on the likelihood of these events.

The axioms imposed on $\{\succsim_A\}_{A \in \mathcal{A}}$ ensure that both π and $\mathbb{N}(\pi)$ can be identified. In addition to SEU axioms and standard axioms relating ex ante and ex post preferences, two main axioms weakening dynamic consistency are imposed. Consider a *minimal unambiguous event* E, i.e., any nonempty $D \subsetneq E$ is ambiguous. The first main axiom, *robust inference*, requires that the DM's ex ante willingness to bet on E is identical to her willingness to bet on E when $D \subsetneq E$ is ruled out. This reflects the DM's cautious attitude when she updates her prior. Since the DM knows the likelihood of unambiguous events but can only guess the likelihood of ambiguous events, when the DM receives new information, she wants her posterior not to differ too much from her benchmark prior on unambiguous events.

The second main axiom, *consistency*, states that every $D \subsetneq E$ is a perfectly dynamically consistent event whenever E is a minimal unambiguous event. As stated earlier when E is unambiguous both E and E^c must be perfectly dynamically consistent. On the other hand, consistency requires that an ambiguous event $D \subsetneq E$ is perfectly dynamically consistent. (This implies that the realization of D^c must lead to a violation of perfect dynamic consistency.) The intuition for this axiom is that when $D \subsetneq E$ is realized, the DM does not learn any new information that can help her make an inference regarding the relative likelihoods of the states within D. To illustrate, consider a DM who is told that an urn contains 25 red (R) balls and 75 blue (B), green (G), or yellow (Y) balls in an unknown proportion. Here, $\{B, G, Y\}$ is a minimal unambiguous event that is known to occur with 0.75 probability, and hence the axiom implies that the DM's preferences are dynamically consistent when, for example, $\{B, G\}$ is realized. This is because the information that $\{B, G\}$ has occurred does not say anything regarding the relative proportion of blue and green balls. Therefore, $f_B \succcurlyeq f_G$ if and only if $f_B \succcurlyeq_{\{B,G\}} f_G$.

1.1.1 Literature Review

This paper lies in the intersection of the literature on non-Bayesian updating and updating under ambiguity. Maximum likelihood updating was introduced by Gilboa and Schmeidler (1993) as a dynamic extension of the maxmin expected utility model. In their model, a DM endowed with a set of priors evaluates acts according to their minimal expected utility, where the minimum is taken over all priors in this set, and the DM performs maximum likelihood updating to revise the set of priors when she receives new information. In Gilboa and Schmeidler (1993), a DM whose behavior is consistent with the subjective expected utility model must follow Bayes' rule. On the other hand, I allow the DM to deviate from Bayesian updating when she faces ambiguity even if she is ambiguity neutral and also show how violations of dynamic consistency can be used to identify the set of priors she considers plausible. The second representation in this paper which has ambiguity averse decision makers is a special case of Gilboa and Schmeidler (1993).

Ortoleva (2012) axiomatizes a novel updating rule, the Hypothesis Testing (HT) model. In his model, the DM follows Bayes' rule for "normal" events but deviates from Bayesian updating when an "unexpected," small probability event occurs. In addition to allowing deviations from Bayesian updating, the HT model also imposes a structure on belief updating when a zero probability event occurs, which is not the case in RML updating. On the other hand, the HT model has two assumptions that are more general than RML updating: (i) the HT model imposes no structure on the set of priors the DM considers, whereas in RML updating every plausible prior must agree with the benchmark prior on unambiguous events, (ii) in the HT model any subjective second-order prior over the set of priors is allowed, while in RML updating it is uniform. In addition, when every state is non-null, the HT model imposes almost no restriction on posteriors, and hence it is significantly more general than RML updating. Due to its generality, the HT model does not have the uniqueness properties of RML updating. In RML updating, both the benchmark prior and the set of plausible priors can be uniquely identified from preferences.

Epstein (2006) and Epstein, Noor, and Sandroni (2008) axiomatize a non-Bayesian updating model where decision makers may be tempted to update their beliefs using a prior different from their original prior. For example, they might be tempted to overreact to new information. In RML updating, decision makers also revise their original prior when they receive new information, but this is not due to temptation but rather due to ambiguity and willingness to make an inference. In addition, the primitive of the analysis is different in these papers.

Zhao (2017) proposes a model that allows DMs to update their beliefs when they receive new information of the form "event A is more likely than event B." In his model, the posterior minimizes Kullback-Leibler (KL) divergence from the prior subject to the constraint that the posterior assigns a higher probability to A than B. In RML updating, the idea is similar as the DM chooses her posterior by minimizing KL divergence from the Bayesian posterior of the benchmark prior subject to the constraint that the new prior must assign the maximal likelihood to the observed event among all plausible priors. Zhao (2018) axiomatizes a non-Bayesian updating model, similarity-based updating, that builds on the representativeness heuristic of Kahneman and Tversky (1972). While both similarity-based updating and RML updating can explain the same phenomena such as the base rate fallacy, the underlying behavioral motivations for these models are completely different.

Many behavioral models in the literature explain non-Bayesian updating by assuming some type of bounded rationality. This includes assuming imperfect memory (Mullainathan, 2002a; Gennaioli and Shleifer, 2010), coarse thinking (Mullainathan, 2002b; Mullainathan, Schwartzstein, and Shleifer, 2008), the use of representativeness heuristic (Kahneman and Tversky, 1972), or incorrect modeling (Barberis, Shleifer, and Vishny, 1998; Rabin and Schrag, 1999) by decision makers. All of these models are non-axiomatic and focus on particular applications.

Since in RML updating the DM is endowed with multiple priors, an alternative natural way to update them is to update each prior according to Bayes' rule. This method is known as full Bayesian updating and was axiomatized by Pires (2002) using the maxmin expected utility model. For ambiguity neutral decision makers, full Bayesian updating rule coincides with Bayesian updating.

While dynamic consistency is usually considered a desirable property, it is frequently violated in the models of belief updating under ambiguity. Epstein and Schneider (2003) retain dynamic consistency by restricting the set of events on which the DM can update her beliefs. Hanany and Klibanoff (2007) characterize dynamically consistent maxmin expected utility preferences without any restriction on the set of conditioning events and show that updated preferences must depend on the initial menu the DM is offered. Siniscalchi (2011) allows deviations from dynamic consistency but assumes that the DM can anticipate her future deviations. Gul and Pesendorfer (2018) impose a weaker version of dynamic consistency which can be interpreted as "not all news can be bad news" and show that neither maximum likelihood updating nor full Bayesian updating satisfies this property. RML updating also treats dynamic consistency as a desirable property by ensuring that the DM is maximally dynamically consistent. That is, any deviation from dynamic consistency is due to the DM's willingness to use new information to make an inference on the set of plausible priors.

Lastly, a few recent papers use ambiguity to explain deviations from Bayesian updating. Baliga, Hanany, and Klibanoff (2013) show that belief polarization can arise when decision makers are ambiguity averse. Fryer, Harms, and Jackson (2018) explain confirmation bias by assuming that when decision makers receive ambiguous signals they interpret it as favorable to their original beliefs. In a social learning experiment, Filippis, Guarino, Jehiel, and Kitagawa (2016) find that decision makers frequently deviate from Bayesian updating when they receive private information that contradicts their original beliefs. Their explanation for this phenomenon assumes multiple priors. All of these papers are non-axiomatic and address specific deviations from Bayesian updating.

The paper proceeds as follows. Section 1.2 introduces the updating rule. In Section 1.3, I take the DM's ex ante and ex post preferences over acts as the primitive and provide a set of behavioral postulates that characterize the updating rule. Section 1.4 illustrates how the model can explain many well-known biases in probabilistic reasoning. In Section 1.5, I extend the model to allow for ambiguity averse preferences. Section 1.6 concludes. Section 1.7 includes all the proofs omitted from the main text.

1.2 Updating Rule

Let Ω be a finite set of states, and denote by $\Delta(\Omega)$ the set of all probability measures on Ω. The collection of all nonempty subsets of Ω is denoted by $\mathcal{A}$. The decision maker's *probability assessment* is characterized by $(\pi, \mathcal{P})$ where $\pi \in \Delta(\Omega)$ is her benchmark prior and $\mathcal{P}$ is a partitioning of Ω that represents the collection of minimal unambiguous events. That is, for any $P \in \mathcal{P}$, the DM assesses that its likelihood is given by $\pi(P)$, and for any nonempty $D \subsetneq P$, the likelihood assigned by π reflects the DM's best guess. Any arbitrary union of the events in $\mathcal{P}$ is unambiguous. A prior is plausible if it agrees with the benchmark prior on unambiguous events. The set of all plausible priors $\mathbb{N}_{\mathcal{P}}(\pi)$ is

$$\mathbb{N}_{\mathcal{P}}(\pi) = \{\pi' \in \Delta(\Omega) |\, \pi'(P) = \pi(P) \text{ for all } P \in \mathcal{P}\}.$$

The set $\mathbb{N}_{\mathcal{P}}(\pi)$ is uniquely defined given the benchmark prior and the collection of minimal unambiguous events.

Suppose the DM learns that event $A \in \mathcal{A}$ is realized. In the model, belief updating can be described by two stages. In the first stage, the DM restricts her attention to the subset of plausible priors that maximize the likelihood that A occurs. Let $\mathbb{N}_{\mathcal{P},A}(\pi)$ denote this set. Formally,

$$\mathbb{N}_{\mathcal{P},A}(\pi) = \underset{\pi' \in \mathbb{N}_{\mathcal{P}}(\pi)}{\arg\max}\, \pi'(A). \tag{1.1}$$

Next, the DM chooses a new benchmark prior from $\mathbb{N}_{\mathcal{P},A}(\pi)$ and updates it using Bayes' rule.

It is required that the new benchmark prior is chosen in a way that its posterior π_A satisfies the following condition:

$$\frac{\pi_A(\omega)}{\pi_A(\omega')} = \frac{\pi(\omega)}{\pi(\omega')} \quad \text{whenever } \omega, \omega' \in A \cap P \text{ for some } P \in \mathcal{P}. \tag{1.2}$$

As shown in Proposition 1.2.1, this is equivalent to requiring that the posterior π_A is the "closest" to the Bayesian posterior of π among all Bayesian posteriors of the priors in $\mathbb{N}_{\mathcal{P},A}(\pi)$, where closeness is defined in terms of Kullback-Leibler divergence.

Given a probability assessment $(\pi, \mathcal{P})$, the posterior π_A is uniquely defined and it is potentially distinct from the Bayesian posterior, which is denoted by $\pi(\cdot|A)$. π_A is interpreted as the DM's benchmark posterior and the new collection of minimal unambiguous events is $\{A \cap P |\, P \in \mathcal{P}\}$. The next lemma illustrates the connection between the posterior π_A and the benchmark prior π.

Lemma 1.2.1. *Let $(\pi, \mathcal{P})$ stand for the DM's probability assessment. For any non-null $A \in \mathcal{A}$ and $\omega \in A$, the posterior π_A obtained via (1.1) and (1.2) satisfies*

$$\pi_A(\omega) = \pi(\omega | A \cap P_\omega) \cdot \pi(P_\omega | \textstyle\bigcup_{P \in \mathcal{P}:\, A \cap P \neq \emptyset} P) \tag{1.3}$$

where P_ω is the member of $\mathcal{P}$ that contains ω.

To understand the formula (1.3), notice that for any $\pi' \in \mathbb{N}_{\mathcal{P},A}(\pi)$, $\pi'(A \cap P) = \pi(P)$ for all $P \in \mathcal{P}$ with $A \cap P \neq \emptyset$. Since π_A is the Bayesian posterior of some $\pi' \in \mathbb{N}_{\mathcal{P},A}(\pi)$, for any $P, P' \in \mathcal{P}$ that have nonempty intersections with A,

$$\frac{\pi_A(P)}{\pi_A(P')} = \frac{\pi'(P|A)}{\pi'(P'|A)} = \frac{\pi'(A \cap P)}{\pi'(A \cap P')} = \frac{\pi(P)}{\pi(P')}.$$

This together with (1.2) show that the posterior π_A satisfies (1.3).

Definition 1.2.1. *Given a probability assessment $(\pi, \mathcal{P})$, the **robust maximum likelihood (RML)** updating rule assigns every non-null event $A \in \mathcal{A}$ the posterior π_A given by (1.3).*

The RML updating rule reflects the DM's awareness of potential inaccuracy of her benchmark prior on ambiguous events, which necessitates a revision of the benchmark prior when new information is received, and her willingness to stay as "close" to her benchmark prior as possible. If all events are unambiguous (i.e., $\mathcal{P}$ is the collection of singletons), RML and Bayesian updating coincide. The next proposition provides an alternative representation for the updating rule.

Proposition 1.2.1. *Let* $(\pi, \mathcal{P})$ *be a probability assessment, and denote by* $\pi(\cdot|A)$ *the Bayesian posterior of* π *when an event* $A \in \mathcal{A}$ *occurs. Then* π_A *is the RML posterior of* π *if and only if*

$$\pi_A = \underset{\pi'_A \in B(\mathbb{N}_{\mathcal{P},A}(\pi))}{\arg\min} D_{KL}(\pi(\cdot|A) \,||\, \pi'_A)$$

where

$$D_{KL}(\pi(\cdot|A) \,||\, \pi'_A) = -\sum_{\omega \in A} \pi(\omega|A) \ln \left(\frac{\pi'_A(\omega)}{\pi(\omega|A)} \right)$$

and $B(\mathbb{N}_{\mathcal{P},A}(\pi))$ *is the set of Bayesian posteriors of the priors in* $\mathbb{N}_{\mathcal{P},A}(\pi)$.

To illustrate the RML updating rule, consider the Ellsberg experiment where the DM is told that an urn contains 30 red (R) balls and 60 blue (B) or green (G) balls in an unknown proportion. Let $(\pi, \mathcal{P})$ stand for the DM's probability assessment and suppose the benchmark prior π assigns equal probability to all colors. For consistency with the information given to the DM, the collection of minimal unambiguous events is $\mathcal{P} = \{\{R\}, \{B, G\}\}$, and the set of plausible priors $\mathbb{N}_{\mathcal{P}}(\pi)$ is

$$\mathbb{N}_{\mathcal{P}}(\pi) = \{\pi' \in \Delta(\{R, B, G\}) | \, \pi'(R) = 1/3\}.$$

Suppose the experimenter draws a ball from the urn and tells the DM that the ball is not green. In this case, the plausible prior that assigns zero probability to green maximizes the likelihood of the observation. Therefore, the RML posterior $\pi_{\{R,B\}}$ is

$$\pi_{\{R,B\}}(R) = 1/3, \quad \pi_{\{R,B\}}(B) = 2/3, \quad \pi_{\{R,B\}}(G) = 0.$$

When the DM is told that the ball is not red, the first stage of RML updating imposes no restriction on the posterior as all plausible priors agree on the event $\{B, G\}$. Hence, in this case the RML posterior $\pi_{\{B,G\}}$ is the same as the Bayesian posterior:

$$\pi_{\{B,G\}}(R) = 0, \quad \pi_{\{B,G\}}(B) = 1/2, \quad \pi_{\{B,G\}}(G) = 1/2.$$

1.3 Representation Theorem

Let X stand for the set of prizes which is assumed to be a convex subset of a metric linear space. For example, X can be the set of monetary outcomes the agent may receive ($X \subseteq \mathbb{R}$) or it can be the set of all lotteries over a finite set of outcomes Z (the classical Anscombe and Aumann (1963) setup). An act assigns a prize to each state of the world. The set of all acts is denoted by $\mathcal{F} = X^{\Omega}$. As usual, constant acts are identified with X. A mixture of two acts is defined statewise: i.e., for any $f, g \in \mathcal{F}$ and $\alpha \in [0,1]$, the act $\alpha f + (1-\alpha)g \in \mathcal{F}$ is defined by $(\alpha f + (1-\alpha)g)(\omega) := \alpha f(\omega) + (1-\alpha)g(\omega)$ for all $\omega \in \Omega$. For any event $A \in \mathcal{A}$ and $f, g \in \mathcal{F}$, $fAg \in \mathcal{F}$ is defined by $(fAg)(\omega) = f(\omega)$ if $\omega \in A$ and $(fAg)(\omega) = g(\omega)$ if $\omega \in A^c$.

I impose axioms on the collection of preferences $\{\succsim_A\}_{A \in \mathcal{A}}$ where $\succsim_A$ reflects the DM's preference over acts when she learns that $A \in \mathcal{A}$ is realized. The DM's preference over acts when she receives no information is $\succsim_\Omega$, which is simply denoted by $\succsim$. For notational simplicity, it is assumed that the DM is indifferent between all acts when the impossible event occurs, i.e., $f \sim_\emptyset g$ for all $f, g \in \mathcal{F}$.

The first three axioms are standard Weak Order, Archimedean, and Independence.

Axiom 1.3.1. *(Weak Order) For any $A \in \mathcal{A}$, $\succsim_A$ is complete and transitive.*

Axiom 1.3.2. *(Archimedean) For any $A \in \mathcal{A}$ and $f, g, h \in \mathcal{F}$ such that $f \succ_A g \succ_A h$, there exist $\alpha, \beta \in (0,1)$ such that $\alpha f + (1-\alpha)h \succ_A g$ and $g \succ_A \beta f + (1-\beta)h$.*

Axiom 1.3.3. *(Independence) For any $A \in \mathcal{A}$, if $f \succ_A g$ and $\alpha \in (0,1]$, then $\alpha f + (1-\alpha)h \succ_A \alpha g + (1-\alpha)h$ for all $h \in \mathcal{F}$.*

Axiom 1.3.4 states that there exist best and worst alternatives and the DM is not indifferent between them. The existence of best and worst alternatives is not necessary for the representation, but it is assumed for the sake of convenience. The assumption that the DM is not indifferent between all alternatives is necessary for the benchmark prior to be identified from preferences.

Axiom 1.3.4. *(Nontriviality) There exist x^* and x_* such that $x^* \succ x_*$ and $x^* \succcurlyeq x \succcurlyeq x_*$ for all $x \in X$.*

The next axiom states that if f assigns a better prize to every state of the world than g does, then f must be preferred to g.

Axiom 1.3.5. *(Monotonicity) For any $A \in \mathcal{A}$, if $f(\omega) \succcurlyeq_A g(\omega)$ for all $\omega \in \Omega$, then $f \succcurlyeq_A g$. If, in addition, f is a constant act and $f(\omega) \succ_A g(\omega)$ for some $\omega \in A$, then $f \succ_A g$.*

Axiom 1.3.5 also requires that if the prize associated with a constant act is replaced in some state with a prize that is strictly worse, the DM considers the new act as strictly inferior. In addition to guaranteeing that the utility function derived from preferences is state independent, Axiom 1.3.5 also ensures that every state is non-null. Notice that by itself this axiom is weaker than strict monotonicity, which requires that if $f(\omega) \succcurlyeq_A g(\omega)$ for all $\omega \in \Omega$ and $f(\omega) \succ_A g(\omega)$ for some $\omega \in A$, then $f \succ_A g$. For example, if the DM evaluates acts according to their worst prize on A, then Axiom 1.3.5 is still satisfied even though strict monotonicity is violated. In the presence of previous axioms, Axiom 1.3.5 and strict monotonicity are equivalent.

The next axiom states that the ranking of two constant acts does not change when new information is received. This is because the prize associated with a constant act is the same regardless of the realized state and the utility of a prize is not affected by new information.

Axiom 1.3.6. *(Constant Act Preference Invariance) For any $A \in \mathcal{A} \setminus \emptyset$ and $x, y \in X$, $x \succcurlyeq y \Leftrightarrow x \succcurlyeq_A y$.*

Axiom 1.3.7 requires that when A is realized the DM must be indifferent between acts that agree on A. This axiom is known as *consequentialism*. In the literature, deviations from consequentialism are usually allowed to accommodate non-expected utility preferences (e.g., Machina,

1989) or to model a decision maker with an imperfect understanding of the state space (e.g., Minardi and Savochkin, 2017). However, the main goal of this paper is to explore non-Bayesian updating when the DM has expected utility preferences and perfect understanding of the state space.

Axiom 1.3.7. *(Consequentialism) For any $A \in \mathcal{A}$, if $f(\omega) = g(\omega)$ for all $\omega \in A$, then $f \sim_A g$.*

If, in addition to Axioms 1.3.1–1.3.7, one also assumes dynamic consistency, then belief updating must be Bayesian (e.g., see Ghirardato, 2002).[5] Dynamic consistency requires that if two acts agree outside an event, then the ranking of these acts should not change when this event occurs. In other words, this says that ex ante optimal plans must be optimal ex post. It can formally be stated as follows.

Dynamic Consistency: For any non-null $A \in \mathcal{A}$ and $f, g \in \mathcal{F}$, $fAg \succsim g \Leftrightarrow f \succsim_A g$.

Axioms 1.3.1–1.3.5 guarantee that the benchmark prior can be uniquely revealed from ex ante preferences as in Anscombe and Aumann (1963). If the DM considers the benchmark prior as the only plausible prior (i.e., no ambiguity), then dynamic consistency is natural. In contrast, if the DM considers multiple priors plausible, it seems natural to revise the benchmark prior when new information arrives. Since the new prior may be distinct from the original benchmark prior, preferences may violate dynamic consistency. However, dynamic consistency should still be satisfied when an unambiguous event is realized. This is because the realization of such an event is not useful in distinguishing between plausible priors as all priors agree on the likelihood of these events, and hence there is no reason for the DM to deviate from her benchmark prior. Therefore, every unambiguous event must be dynamically consistent defined as below.

Definition 1.3.1. *$A \in \mathcal{A}$ is **dynamically consistent** if for any $f, g \in \mathcal{F}$, $fAg \succsim g \Leftrightarrow f \succsim_A g$.*[6]

[5]The connection between dynamic consistency and Bayesian updating is very general as shown in Epstein and Le Breton (1993). They show that if conditional preferences are derived in a way to ensure dynamic consistency and both ex ante and conditional preferences are "based on beliefs" (i.e., an event A is considered to be more likely than B if the DM prefers to bet on A rather than B), then standard axioms (Savage axioms except the Sure Thing Principle) guarantee that there exists a unique prior that represents beliefs and conditional beliefs are obtained using Bayes' rule.

[6]Notice that dynamic consistency is a feature of preferences not events. However, I use this terminology for the sake of brevity.

Suppose the analyst observes that the DM's preferences are dynamically consistent upon realization of an event. Can the analyst conclude that the DM considers this event as unambiguous? I provide an example which shows that this conclusion may not be accurate and then define a stronger version of dynamic consistency that captures unambiguous events.

Example 1.3.1. *Consider a DM who is told that an urn contains 50 red or blue balls and 50 green or yellow balls in unknown proportions. Let* $\Omega = \{R, B, G, Y\}$ *where* R, B, G, *and* Y *stand for states when the ball drawn from an urn is red, blue, green, and yellow, respectively. The set of plausible priors is*

$$\{\pi' \in \Delta(\{R, B, G, Y\})|\, \pi'(R) + \pi'(B) = \pi'(G) + \pi'(Y) = 1/2\}.$$

When no further information is given, the DM may choose her benchmark prior as the one which assigns equal probability to all colors. Now suppose that a ball is drawn from the urn and the DM is told that the ball is either blue or green. This information does not favor either blue or green relative to the original information. Hence, it makes sense to assume that the benchmark posterior also assigns equal probability to blue and green. But then the event $\{B, G\}$ *is dynamically consistent. On the other hand, given the set of plausible priors, it is not possible to tell the exact probability that* $\{B, G\}$ *occurs.*

In Example 1.3.1, even though $\{B, G\}$ is a dynamically consistent event, it is still possible that $\{B, G\}$ is ambiguous. Consider two bets $f_B = (0, 100, 0, 0)$ and $f_G = (0, 0, 100, 0)$. If the DM's benchmark prior and posterior are as in the example, it must be that $f_B \sim f_G$ and $f_B \sim_{\{B,G\}} f_G$. Now suppose before learning that the ball drawn from the urn is either blue or green, the DM first learns that the ball is not yellow. The information that the ball is not yellow may suggest that the number of green balls in the urn is greater than the number of yellow balls. Since this information does not say anything regarding the relative proportion of red and blue balls, there is no reason for the DM to deviate from her original evaluation of the relative likelihood of R and B. But then the DM strictly prefers f_G over f_B when she learns that the ball is not yellow. Hence, $f_G \succ_{\{R,B,G\}} f_B$

even though $f_B \sim_{\{B,G\}} f_G$ and f_B and f_G agree on $\{R\}$. This would not be expected if $\{B, G\}$ was unambiguous.

This example motivates a new definition that captures unambiguous events via a stronger version of dynamic consistency

Definition 1.3.2. *$A \in \mathcal{A}$ is **perfectly dynamically consistent** if for any event $B \supseteq A$ and $f, g \in \mathcal{F}$,*

$$fAg \succsim_B g \Leftrightarrow f \succsim_A g.$$

For a Bayesian decision maker, every event should be perfectly dynamically consistent. Indeed, perfect dynamic consistency is implicitly assumed in the previous characterizations of Bayesian updating as $\{\succsim_A\}_{A\in\mathcal{A}}$ satisfies dynamic consistency only if every non-null event is perfectly dynamically consistent.[7]

Example 1.3.1 illustrates that when $\{\succsim_A\}_{A\in\mathcal{A}}$ does not satisfy dynamic consistency, there may be events that are dynamically consistent but not perfectly dynamically consistent. Since unambiguous events are expected to be perfectly dynamically consistent and it is possible to find a violation of perfect dynamic consistency for ambiguous events as in Example 1.3.1, an event is defined to be unambiguous if the event as well as its complement are perfectly dynamically consistent. The reason for requiring the complement to be perfectly dynamically consistent comes from the observation that the complement of an unambiguous event must be unambiguous.

Definition 1.3.3. *E is an **unambiguous event** if both E and E^c are perfectly dynamically consistent. The collection of all unambiguous events is denoted by $\mathcal{E}$. An event that does not belong to $\mathcal{E}$ is an **ambiguous event**.*

The next axiom ensures that the collection of unambiguous events form an *algebra*. A collection of events $\mathcal{E}$ is an algebra over Ω if (i) $\Omega \in \mathcal{E}$, (ii) $E \in \mathcal{E}$ implies $E^c \in \mathcal{E}$, and (iii) $E, E' \in \mathcal{E}$ implies $E \cap E' \in \mathcal{E}$.

[7] To see this, suppose $\{\succsim_A\}_{A\in\mathcal{A}}$ satisfies dynamic consistency. Let $A \in \mathcal{A}$, $B \supseteq A$ and $f, g \in \mathcal{F}$ be given. Dynamic consistency implies that $fAg \succsim g \Leftrightarrow f \succsim_A g$. On the other hand, since $B \supseteq A$, $fAg = (fAg)Bg$. Hence, dynamic consistency also implies that $fAg \succsim g \Leftrightarrow (fAg)Bg \succsim g \Leftrightarrow fAg \succsim_B g$. Therefore, A is perfectly dynamically consistent.

Axiom 1.3.8. *(Algebra of Unambiguous Events) If $E, E' \in \mathcal{E}$, then $E \cap E' \in \mathcal{E}$.*

Intuitively, this axiom requires the following. Suppose events E, E^c, E', and E'^c are perfectly dynamically consistent so that both E and E' are unambiguous. Let f and g be two acts which agree outside $E \cap E'$. By the definition of perfect dynamic consistency, ex ante preference between f and g must be preserved when the DM learns either $B \supseteq E$ or $B' \supseteq E'$. But then it makes sense to assume that ex ante preference between f and g must still be preserved when the DM learns B and B' simultaneously. Therefore, perfectly dynamically consistent events are expected to be closed under intersection.

Since $\mathcal{E}$ is an algebra, there exists a unique partitioning of the state space that generates $\mathcal{E}$. A partition $\mathcal{P}$ of Ω generates the algebra $\mathcal{E}$ if $E \in \mathcal{E} \Leftrightarrow$ there exist $P_1, \ldots, P_k \in \mathcal{P}$ such that $P_1 \cup \cdots \cup P_k = E$. Let $\mathcal{P}_{\mathcal{E}}$ denote the partition that generates $\mathcal{E}$. The members of $\mathcal{P}_{\mathcal{E}}$ are minimal unambiguous events, i.e., any nonempty $D \subsetneq P$ where $P \in \mathcal{P}_{\mathcal{E}}$ is ambiguous. The next two axioms rely on $\mathcal{P}_{\mathcal{E}}$.

The following definitions will be useful for the statement of the next axiom.

Definition 1.3.4. *1. For any event A, **a bet on** A is an act f_A that yields the best prize on A and the worst prize outside A, i.e., $f_A = x^* A x_*$.*

*2. For any $f \in \mathcal{F}$ and $A \in \mathcal{A}$, **a certainty equivalent of** f **given** A is a sure outcome $c_A(f) \in X$ such that $f \sim_A c_A(f)$.*[8]

Since the DM knows the likelihood of unambiguous events but can only guess the likelihood of ambiguous events, she may want her posterior not to differ too much from her benchmark prior on unambiguous events. The next axiom, *robust inference*, reflects this cautious attitude when the DM updates her benchmark prior. Consider a minimal unambiguous event $P \in \mathcal{P}_{\mathcal{E}}$ and let f_P denote a bet on P. Suppose A is realized, and hence $A \cap P$ is a new minimal unambiguous event. Robust inference requires that the DM's willingness to bet on P is not affected when $D \subsetneq A \cap P$ is ruled out. That is, $c_A(f_P) \sim c_{A \setminus D}(f_P)$. In other words, since D is a proper subset of a minimal

[8]The axioms stated so far and the assumption that X is convex guarantee that every act has a certainty equivalent.

unambiguous event, when it is ruled out, the DM's considers the plausibility that it was a null event in the first place.

Axiom 1.3.9. *(Robust Inference) For any $A \in \mathcal{A}$ and $D \subsetneq A \cap P$ where $P \in \mathcal{P}_{\mathcal{E}}$,*

$$c_A(f_P) \sim c_{A \setminus D}(f_P).$$

In general, it is desirable if the DM's preferences are dynamically consistent unless there is a justifiable reason for deviation. The next axiom, *consistency*, requires that every $D \subsetneq P$, where P is a minimal unambiguous event, is perfectly dynamically consistent. Since an event E is unambiguous when both E and E^c are perfectly dynamically consistent and consistency requires an ambiguous event $D \subsetneq P$ to be perfectly dynamically consistent, the implication of the axiom is that D^c is not perfectly dynamically consistent.

Axiom 1.3.10. *(Consistency) Every $D \subsetneq P$ where $P \in \mathcal{P}_{\mathcal{E}}$ is perfectly dynamically consistent.*

Intuitively, when $D \subsetneq P$ is realized, the DM does not learn any information that can help her make an inference regarding the relative likelihoods of the states within D, and hence there is no reason for the DM to deviate from her original evaluation. To illustrate, suppose the DM is told that an urn contains 25 red (R) balls and 75 blue (B), green (G), or yellow (Y) balls in an unknown proportion. Since $\{B, G, Y\}$ is a minimal unambiguous event that is known to occur with 0.75 probability, consistency requires that the DM's preferences are dynamically consistent when, for example, $\{B, G\}$ is realized. This is because the information that $\{B, G\}$ is realized does not say anything regarding the relative proportion of blue and green balls, and hence there is no justification for deviation from the benchmark prior. Therefore, if the DM is ex ante indifferent between betting on blue and betting on green, she should remain indifferent when she learns that the ball drawn from the urn is either blue or green.

The next theorem provides a characterization result for the RML updating model.

Theorem 1.3.1. *The collection of preferences $\{\succsim_A\}_{A \in \mathcal{A}}$ satisfies Weak Order, Archimedean, Independence, Nontriviality, Mononotonicity, Constant Act Preference Invariance, Consequentialism,* *Algebra of Unambiguous Events, Robust Inference, and Consistency if and only if there exist a non-constant, affine utility function $u : X \to \mathbb{R}$ with $u(X) = [u(x_*), u(x^*)]$ and a probability assessment $(\pi, \mathcal{P})$, where π has full support on Ω, such that for any $A \in \mathcal{A}$,*

$$f \succcurlyeq_A g \Leftrightarrow \sum_{\omega\in\Omega} \pi_A(\omega)u(f(\omega)) \geq \sum_{\omega\in\Omega} \pi_A(\omega)u(g(\omega)) \tag{1.4}$$

and π_A is the RML posterior of π. Moreover, u is unique up to a positive affine transformation, π_A is unique for all $A \in \mathcal{A} \setminus \emptyset$, and $\mathcal{P}$ is uniquely revealed as $\mathcal{P}_{\mathcal{E}}$ unless $\mathcal{P} = \{\Omega, \emptyset\}$.

1.3.1 Sketch of the Proof

While showing the necessity of Axioms 1.3.1–1.3.7 is standard, the necessity of Axioms 1.3.8–1.3.10 is not trivial. The key step in the proof is showing that if $\{\succcurlyeq_A\}_{A\in\mathcal{A}}$ can be represented by (1.4), then the collection of minimal unambiguous events $\mathcal{P}_{\mathcal{E}}$ that is derived from preferences is exactly $\mathcal{P}$ unless $\mathcal{P} = \{\Omega, \emptyset\}$. This is achieved by showing that an event belongs to the algebra generated by $\mathcal{P}$ if and only if both this event and its complement satisfy perfect dynamic consistency. Once this can be shown, Axioms 1.3.8–1.3.10 directly follow from the representation.

The proof of the claim that when E belongs to the algebra generated by $\mathcal{P}$ both E and E^c are perfectly dynamically consistent can be done using standard arguments. To prove the opposite, suppose E does not belong to the algebra generated by $\mathcal{P}$. There are two cases to consider. As an illustration, consider the case when $E \subsetneq P$ for some $P \in \mathcal{P}$. Since Axiom 1.3.10 is satisfied only if E is perfectly dynamically consistent, it needs to be shown that E^c is not perfectly dynamically consistent. Consider bets on P and $P \setminus E$, i.e., $f_P = x^*Px_*$ and $f_{P\setminus E} = x_*E(x^*Px_*)$. Using the representation (1.4), it is possible to find $\bar{z}, z \in X$ such that

$$f_P = \begin{pmatrix} x^* & \text{if } \omega \in E \\ x^* & \text{if } \omega \in P \setminus E \\ x_* & \text{if } \omega \in \Omega \setminus P \end{pmatrix} \sim \bar{z} \quad \text{and} \quad f_{P\setminus E} = \begin{pmatrix} x_* & \text{if } \omega \in E \\ x^* & \text{if } \omega \in P \setminus E \\ x_* & \text{if } \omega \in \Omega \setminus P \end{pmatrix} \sim \begin{pmatrix} x_* & \text{if } \omega \in E \\ z & \text{if } \omega \in P \setminus E \\ z & \text{if } \omega \in \Omega \setminus P \end{pmatrix}$$

where $\bar{z} \succ z$. Now suppose E^c is realized. From the representation, $\bar{z} \succ_{E^c} z \sim_{E^c} x_* E z$. Since π_{E^c} the RML posterior of π, $\bar{z} \sim_{E^c} f_P \sim_{E^c} f_{P \setminus E}$. But then, $f_{P \setminus E} = x_* E(x^* P x_*) \succ_{E^c} x_* E z$, in violation of dynamic consistency. Therefore, E^c is not perfectly dynamically consistent. This shows that E is an ambiguous event, i.e., $E \notin \mathcal{E}$. The case when $E \cap P \neq \emptyset$ and $E \cap P' \neq \emptyset$ for at least two distinct $P, P' \in \mathcal{P}$ is similar.

To prove sufficiency, first observe that Axioms 1.3.1–1.3.5 yield an SEU representation for each $A \in \mathcal{A}$ as in Anscombe and Aumann (1963). Moreover, Axiom 1.3.6 guarantees that the same utility function can be used for all $\succcurlyeq_A$, and Axioms 1.3.5 and 1.3.7 guarantee that π_A has full support on A and $\pi_A(A^c) = 0$. Therefore, it only needs to be shown that each π_A is the RML posterior of π. Let $\mathcal{E}$ be given by Definition 1.3.3, and $\mathcal{P}_{\mathcal{E}}$ is the partition that generates $\mathcal{E}$. The DM's probability assessment is $(\pi, \mathcal{P}_{\mathcal{E}})$. Since preferences are dynamically consistent on $E \in \mathcal{E}$, standard arguments show that updating is Bayesian when E is realized, consistent with RML updating.

Consider an event $A \notin \mathcal{E}$. I construct an unambiguous event $B \in \mathcal{E}$ such that $B \supseteq A$ and B is the smallest such event with respect to set inclusion. To construct B, let $\mathcal{P}_{\mathcal{E}} = \{P_1, \dots, P_n\}$ and consider $J \subseteq \{1, \dots, n\}$ such that $P_j \cap A \neq \emptyset$ for all $j \in J$. The event B is given by $B = \cup_{j \in J} P_j$. Since $B \in \mathcal{E}$, according to the previous paragraph, the DM performs Bayesian updating when B is realized. Hence, for any $j, j' \in J$,

$$\frac{\pi_B(P_j)}{\pi_B(P_{j'})} = \frac{\pi(P_j)}{\pi(P_{j'})}.$$

On the other hand, Axiom 1.3.9 implies that $\pi_A(P_j) = \pi_{A \cup P_j}(P_j)$ for all $j \in J$. Therefore, iterative application of Axiom 1.3.9 yields $\pi_A(P_j) = \pi_B(P_j)$ for all $j \in J$, which implies

$$\frac{\pi_A(P_j)}{\pi_A(P_{j'})} = \frac{\pi(P_j)}{\pi(P_{j'})}.$$

Consider an event $A \cap P_j$ where $j \in J$. By Axiom 1.3.10, $A \cap P_j$ is perfectly dynamically consistent, and hence standard arguments guarantee that $\pi_{A \cap P_j}$ is the Bayesian posterior of π.

Moreover, perfect dynamic consistency also ensures that $\pi_{A\cap P_j}(\omega) = \pi_A(\omega|P_j)$ for all $\omega \in A\cap P_j$. Therefore, for any $\omega, \omega' \in A \cap P_j$,

$$\frac{\pi_A(\omega)}{\pi_A(\omega')} = \frac{\pi_{A\cap P_j}(\omega)}{\pi_{A\cap P_j}(\omega')} = \frac{\pi(\omega)}{\pi(\omega')}.$$

This together with the conclusion of the previous paragraph and Lemma 1.2.1 show that π_A is the RML posterior of π, concluding the proof of sufficiency.

Lastly, the uniqueness of u up to a positive affine transformation and the uniqueness of π_A for each $A \in \mathcal{A}$ are standard results. The uniqueness of $\mathcal{P}$ is implied by the proof of necessity where the equivalence of $\mathcal{P}$ and $\mathcal{P}_{\mathcal{E}}$ is shown.

1.4 Applications

In this section, I show how the RML updating rule can help explain commonly observed biases in probabilistic reasoning. While all the examples in this section only use the first stage of RML updating, in more realistic examples with a larger state space the first stage of RML updating by itself will in general not result in a unique posterior, and hence the second stage of RML updating will be needed to make meaningful predictions.

Let $\Omega \equiv S \times \Sigma$ where $S = \{s_1, s_2\}$ is the set of payoff-relevant states and $\Sigma = \{\sigma_1, \sigma_2\}$ is the set of possible signals. The DM's benchmark prior π is represented by two parameters (μ, α) where $\mu > 1/2$ is the probability that the payoff relevant state is s_1 and $\alpha > 1/2$ denotes the probability that the DM receives signal σ_i when the payoff-relevant state is s_i. $\mu > 1/2$ reflects the DM's initial evaluation that s_1 is more likely.

	σ_1	σ_2
s_1	ω_{11}	ω_{12}
s_2	ω_{21}	ω_{22}

State Space Ω

	σ_1	σ_2
s_1	$\mu\alpha$	$\mu(1-\alpha)$
s_2	$(1-\mu)(1-\alpha)$	$(1-\mu)\alpha$

Benchmark Prior π

1.4.1 Confirmation Bias

I now revisit the confirmation bias phenomenon illustrated in the introduction. The DM who displays confirmation bias interprets contrary evidence as supportive of her original beliefs (Rabin and Schrag, 1999). That is, when the DM observes σ_2, she finds s_1 to be at least as likely as before. Formally,

$$\pi_{\text{conf. bias}}(s_1|\sigma_2) \geq \mu > \frac{\mu - \mu\alpha}{\mu + \alpha - 2\mu\alpha} = \pi(s_1|\sigma_2).$$

Confirmation bias is frequently reported in experiments (e.g., see Lord, Ross, and Lepper, 1979; Darley and Gross, 1983). Here, I provide an explanation for confirmation bias that uses RML updating (see Minardi and Savochkin, 2017; Fryer, Harms, and Jackson, 2018, for other recent models that can accommodate confirmation bias).

Suppose the DM finds it plausible that there is a "bias" in the information source that is potentially unfavorable towards the state she originally finds more likely. Even though such a DM unambiguously knows the probability that the payoff-relevant state is s_1, the event that σ_i occurs when s_1 is the payoff-relevant state is ambiguous. Therefore, when the DM observes σ_2, she revises her benchmark prior to account for the plausibility that, due to the bias, σ_2 might be more likely than σ_1 when the payoff-relevant state is s_1. Notice that the DM might find the existence of a bias plausible even though it is not her benchmark belief. Once the bias is seen as plausible, the DM is endowed with multiple priors and uses signal realizations to distinguish between plausible priors.

To formalize the intuition, let $\mathcal{P}$ stand for the set of minimal unambiguous events in this example, which is given by

$$\mathcal{P} = \{\{\omega_{11}, \omega_{12}\}, \{\omega_{21}\}, \{\omega_{22}\}\}.$$

In RML updating, the DM uses the maximum likelihood method to make an inference regarding the direction of the bias. Given the benchmark prior and the set of minimal unambiguous events, the RML posterior π_{σ_2} is

$$\pi_{\sigma_2}(s_1) = \frac{\mu}{\mu + \alpha - \mu\alpha} \quad \text{and} \quad \pi_{\sigma_2}(s_2) = \frac{\alpha - \mu\alpha}{\mu + \alpha - \mu\alpha}$$

Hence, the DM performing RML updating believes that s_1 is strictly more likely than before when she observes σ_2, consistent with confirmation bias.

1.4.2 Other Behavioral Biases

I consider three other commonly observed deviations from Bayesian updating: base rate neglect, conservatism, and overconfidence.

Base Rate Neglect: In a series of experiments, Kahneman and Tversky (1973) and Bar-Hillel (1980) show that decision makers tend to ignore the base rate μ in their predictions. In the well-known "cab problem," DMs are told that there are two cab companies, Blue and Green, one of which has been involved in a hit-and-run accident. The proportion of Blue cabs in the city is 85%, and the cab involved in the accident was identified as Green by a witness who is accurate 80% of the time. When DMs are asked to predict the probability that the car involved in the accident is Green, the median and modal response is 0.8, much higher than the Bayesian posterior (≈ 0.41).

To see how RML updating can explain this phenomenon, imagine that the DM has full confidence in the likelihood information α but does not have full confidence in the base rate μ. Even when the DM does not have full confidence in the base rate, the event that consists of states in which she gets "correct" signals is still unambiguous and known to occur with $\alpha = 0.8$ probability. Similarly, the event that corresponds to states in which she gets "wrong" signals is unambiguously assigned $1 - \alpha = 0.2$ probability. Given this set of minimal unambiguous events, the RML posterior is exactly equal to the median response in the cab problem (see the figure below).

Conservatism: DMs display conservatism bias when they overweight the base rate and underweight the likelihood information (see Edwards, 1968, for the classical experimental findings). RML updating results in conservatism bias when decision makers have full confidence in the base rate information but not in the likelihood information. This is exactly the mirror image of the base rate neglect phenomenon.

Overconfidence: Decision makers who treat their private information as more precise than it actually is are described as overconfident (see Odean, 1998, for a review of psychology literature on overconfidence and its implications for asset markets). Suppose s_1 = good market, s_2 = bad market, σ_1 = good jobs report, and σ_2 = bad jobs report. Overconfident investors tend to over-invest when they observe good jobs report and under-invest when they observe bad jobs report. RML updating results in overconfidence when the DM has full confidence in the likelihood information but is not completely sure whether the "correct" signal is more likely when the state is s_1 or s_2.

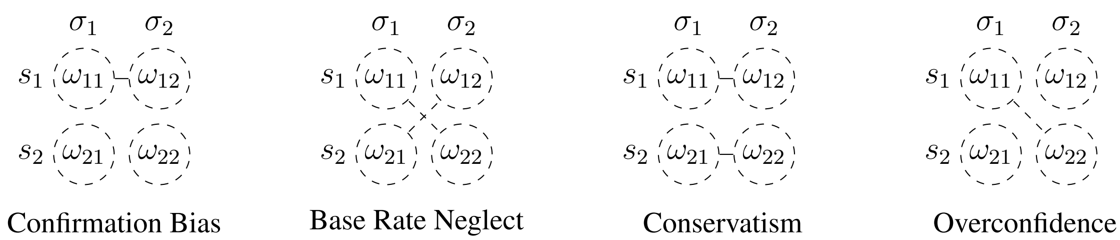

Figure 1.1: Behavioral biases. This figure illustrates partitions used to explain each behavioral bias. States connected by a line belong to the same partition element.

1.5 Ambiguity Averse Preferences

In Section 1.3, ambiguity is reflected in the DM's belief updating even though the DM's preferences display neutral attitude towards ambiguity. In this section, I consider an ambiguity averse DM whose preferences are consistent with the maxmin expected utility model of Gilboa and Schmeidler (1989).

An ambiguity averse DM is expected to satisfy all the axioms that characterize the subjective expected utility model in Section 1.3 except independence and consistency. An ambiguity averse DM may not satisfy independence due to strict preference for randomization, which may arise as randomization potentially limits the exposure to ambiguity. In the Ellsberg example, the DM may be indifferent between betting on blue (f_B) and betting on green (f_G) but may strictly prefer the 50-50 randomization of these bets, in violation of independence (see below). This is because the 50-50 randomization between f_B and f_G gives the DM the same monetary outcome regardless of whether the ball drawn from the urn is blue or green, and hence it can be seen as a perfect hedge against ambiguity.

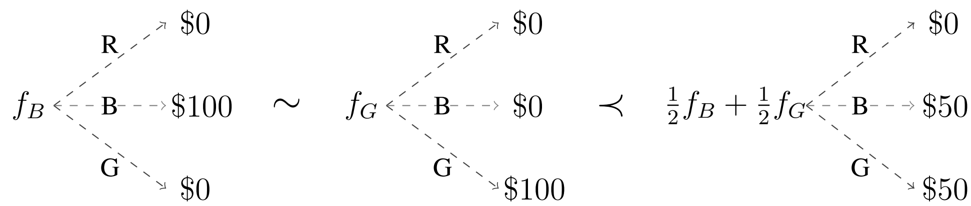

Figure 1.2: Preference for randomization due to ambiguity aversion.

To illustrate why an ambiguity averse DM may not satisfy consistency, recall the example where the DM is told that an urn contains 25 red (R) balls and 75 blue (B), green (G), or yellow (Y) balls in an unknown proportion. Here, $\{B, G, Y\}$ is a minimal unambiguous events, and hence consistency implies that $\{B, G\}$ is dynamically consistent. Let f_1 be the act that yields \$100 if the ball drawn from the urn is blue, and \$0 otherwise. Let f_2 be the act that yields \$25 if the ball drawn from the urn is either blue or green, and \$0 otherwise. When the DM learns that $\{B, G\}$ is realized, she may have a strict preference for f_2 over f_1 as f_2 perfectly hedges against ambiguity but f_1 does not. Should this DM have a strict preference for f_2 over f_1 ex ante as required by consistency? This is not clear because ex ante f_2 is not a perfect hedge against ambiguity and it has much lower expected value than f_1 for many plausible priors.

In addition to the axioms in Section 1.3 except independence and consistency, I impose two new axioms on preferences that characterize RML updating for ambiguity averse DMs. Let $\mathcal{E}$ stand for the collection of unambiguous events as in Definition 1.3.3, and $\mathcal{P}_\mathcal{E}$ is the collection of minimal unambiguous events. Acts that are constant on minimal unambiguous events are unambiguous acts.

Definition 1.5.1. *$f \in \mathcal{F}$ is an **unambiguous act** if $f(\omega) = f(\omega')$ whenever $\omega, \omega' \in P$ for some $P \in \mathcal{P}_\mathcal{E}$. The set of all unambiguous acts is denoted by $\mathcal{F}^{ua} \subseteq \mathcal{F}$.*

Axiom 1.5.1 imposes independence on the set of all unambiguous acts. Since unambiguous acts have no exposure to ambiguity, strict preference for randomization between unambiguous acts cannot be justified by ambiguity aversion.

Axiom 1.5.1. *(Weak Independence) For any $A \in \mathcal{A}$, $f, g, h \in \mathcal{F}^{ua}$ and $\alpha \in (0, 1]$, $f \succ_A g$ implies $\alpha f + (1-\alpha)h \succ_A \alpha g + (1-\alpha)h$.*

Weak independence is consistent with both ambiguity averse and ambiguity loving attitude. The next axiom imposes that the DM is ambiguity averse.

Axiom 1.5.2. *(Ambiguity Aversion) For any $A \in \mathcal{A}$, $D \subsetneq A \cap P$ where $P \in \mathcal{P}_{\mathcal{E}}$, and $f \in \mathcal{F}$, $x^* D f \sim_A f$.*

Consider a minimal unambiguous event P and suppose A is realized. After this realization, $A \cap P$ is a minimal unambiguous event, and hence $D \subsetneq A \cap P$ is ambiguous. Axiom 1.5.2 requires that the DM is indifferent between an act f and an act which agrees with f outside D and yields the best prize on D. This is an extreme attitude that completely disregards that in the second act the DM receives the best prize when D occurs. This is due to two assumptions: (i) in the model, every $D \subsetneq A \cap P$ is treated as maximally ambiguous, (ii) given the set of plausible priors, the DM evaluates acts according to their worst case utility as in Gilboa and Schmeidler (1989).

The next theorem provides a characterization for RML updating for ambiguity averse DMs.

Theorem 1.5.1. *The collection of preferences $\{\succsim_A\}_{A \in \mathcal{A}}$ satisfies Weak Order, Archimedean, Weak Independence, Nontriviality, Monotonicity, Constant Act Preference Invariance, Consequentialism, Algebra of Unambiguous Events, Robust Inference, and Ambiguity Aversion if and only if there exist a non-constant, affine utility function $u : X \to \mathbb{R}$ with $u(X) = [u(x_*), u(x^*)]$, and a probability assessment $(\pi, \mathcal{P})$, where π has full support on $\mathcal{P}$, such that for any $A \in \mathcal{A}$,*

$$f \succsim_A g \Leftrightarrow \min_{\pi_A \in B(\mathbb{N}_{\mathcal{P},A}(\pi))} \sum_{\omega \in \Omega} \pi_A(\omega) u(f(\omega)) \geq \min_{\pi_A \in B(\mathbb{N}_{\mathcal{P},A}(\pi))} \sum_{\omega \in \Omega} \pi_A(\omega) u(g(\omega)) \tag{1.5}$$

where $B(\mathbb{N}_{\mathcal{P},A}(\pi))$ is the set of Bayesian posteriors of the priors in $\mathbb{N}_{\mathcal{P},A}(\pi)$. Moreover, u is unique up to a positive affine transformation, $\mathcal{P}$ is uniquely revealed as $\mathcal{P}_{\mathcal{E}}$, and the set $\mathbb{N}_{\mathcal{P},A}(\pi)$ is unique for all $A \in \mathcal{A} \setminus \emptyset$.

When the DM's preferences are consistent with the maxmin expected utility model, the benchmark prior can no longer be identified from preferences. In fact, the only role of the benchmark prior in Theorem 1.5.1 is to define the likelihood of unambiguous events. Since the DM uses the worst case scenario to evaluate acts, even if she has a guess for the likelihood of ambiguous events this will not be reflected in her preferences. This distinguishes Theorem 1.5.1 from Theorem 1.3.1 where the benchmark prior can be revealed from ex ante preferences. Because of the limited role the benchmark prior plays in Theorem 1.5.1, with maxmin expected utility preferences RML updating coincides with the maximum likelihood updating rule of Gilboa and Schmeidler (1993).

Recall that an event E is defined to be unambiguous if and only if both E and E^c are perfectly dynamically consistent, which holds for both ambiguity neutral and ambiguity averse DMs. Example 1.3.1 shows that an event may fail to be perfectly dynamically consistent even when it is dynamically consistent. However, if $\{\succsim_A\}_{A\in\mathcal{A}}$ satisfies the axioms in Theorem 1.5.1, then every dynamically consistent event is perfectly dynamically consistent. Therefore, if the DM is ambiguity averse, it is possible to identify unambiguous events via dynamic consistency.

Lemma 1.5.1. *If the collection of preferences $\{\succsim_A\}_{A\in\mathcal{A}}$ can be represented by (1.5), then every dynamically consistent event is perfectly dynamically consistent, and hence an event E is unambiguous if and only if both E and E^c are dynamically consistent.*

1.6 Conclusion

Many real life economic problems involve ambiguity. In this paper, it is argued that departing from Bayesian updating is natural when one faces ambiguity. I axiomatize a non-Bayesian updating model, robust maximum likelihood (RML) updating, where the DM's probability assessment can be represented by a benchmark prior, which reflects the DM's initial best guess, and a set of priors the DM considers plausible. The DM responds to new information by revising the benchmark prior via the maximum likelihood principle in a way that ensures maximally dynamically consistent behavior, and updates the new prior using Bayes' rule. I show that RML updating can accommodate many commonly observed deviations from Bayesian updating.

I take the DM's preferences over acts before and after the arrival of new information as the primitive of the analysis. In addition to standard axioms, the two main axioms imposed on preferences are robust inference and consistency. Robust inference requires that when a proper subset of a minimal unambiguous event is ruled out, the DM's willingness to bet on this minimal unambiguous event is not affected. This reflects the DM's cautious attitude when she updates her benchmark prior. Consistency states that every proper subset of a minimal unambiguous event is perfectly dynamically consistent. This reflects the intuition that when such an event is realized, the DM does not learn any new information that can help her make an inference regarding the relative likelihoods of the states within this event. I show that if the DM satisfies these axioms, both the benchmark prior and the set of plausible priors are uniquely identified from preferences.

1.7 Appendix

Proof of Theorem 1.3.1

Necessity

The necessity of Axioms 1.3.1–1.3.5 is standard. The necessity of constant act preference invariance follows from the observation that the same utility function is used for all $\succsim_A$ in the representation. Consequentialism is necessary as the RML posterior satisfies $\pi_A(\omega) > 0$ only if $\omega \in A$.

Let $(\pi, \mathcal{P}, u)$ be a representation of $\{\succsim_A\}_{A \in \mathcal{A}}$ given by Theorem 1.3.1. To prove the necessity of Axioms 1.3.8–1.3.10, I show that the set of unambiguous events, denoted by $\mathcal{E}$, is equal to $\sigma(\mathcal{P})$, the algebra generated by the partition $\mathcal{P}$, as long as $\mathcal{P}$ is not degenerate (i.e., $\mathcal{P} \neq \{\Omega, \emptyset\}$). This is proved in Lemma 1.7.3 after two preliminary observations. The next lemma shows that if $A \subseteq B$

and π_A and π_B are the RML posteriors of π, then π_A is the RML posterior of π_B where the partition of B is given by $\{B \cap P | \, P \in \mathcal{P}\}$.

Lemma 1.7.1. *Let $(\pi, \mathcal{P})$ represent the probability assessment and suppose for any non-null $A \in \mathcal{A}$, π_A is the RML posterior of π. Then for any non-null $A \subseteq B$, π_A is the RML posterior of π_B where the partition of B is given by $\{B \cap P | \, P \in \mathcal{P}\}$.*

Proof. Notice that since for any $P \in \mathcal{P}$, $A \cap (B \cap P) = A \cap P$, it suffices to show that for any P, P' with $A \cap P \neq \emptyset$ and $A \cap P' \neq \emptyset$,

$$\frac{\pi_A(P)}{\pi_A(P')} = \frac{\pi_B(P)}{\pi_B(P')}$$

and for any $\omega, \omega' \in A \cap P$ where $P \in \mathcal{P}$,

$$\frac{\pi_A(\omega)}{\pi_A(\omega')} = \frac{\pi_B(\omega)}{\pi_B(\omega')}.$$

The second equation follows from the observation that since π_A and π_B are the RML posteriors of π and $A \subseteq B$, the above ratio is equal to $\frac{\pi(\omega)}{\pi(\omega')}$. Similarly, the first equation follows from the observation that $B \cap P \neq \emptyset$ whenever $A \cap P \neq \emptyset$, and hence the above ratio is equal to $\frac{\pi(P)}{\pi(P')}$. $\square$

Lemma 1.7.2. *Let $E \subsetneq P$ for some $P \in \mathcal{P}$ and $f_{P \setminus E} = x_* E(x^* P x_*)$. Then for any $A \supseteq P$, there exists $z_A \in X$ such that $f_{P \setminus E} \sim_A x_* E z_A$.*

Proof. Let z_A be defined by

$$z_A = \frac{\pi_A(P \setminus E)}{1 - \pi_A(E)} x^* + \frac{1 - \pi_A(P)}{1 - \pi_A(E)} x_*.$$

By assumption, $A \supsetneq E$ and $\pi_A(\omega) > 0$ for each $\omega \in A$. Hence, $\pi_A(E) < 1$. Since X is convex, $z_A \in X$. Moreover, since u is affine,

$$u(z_A) = \frac{\pi_A(P \setminus E)}{1 - \pi_A(E)} u(x^*) + \frac{1 - \pi_A(P)}{1 - \pi_A(E)} u(x_*).$$

Let U_A be a representation of $\succcurlyeq_A$ given by Theorem 1.3.1. By the representation,

$$\begin{aligned}U_A(f_{P\setminus E}) &= \pi_A(P\setminus E)u(x^*) + (1-\pi_A(P\setminus E))u(x_*)\\ &= \pi_A(P\setminus E)u(x^*) + (1-\pi_A(P))u(x_*) + \pi_A(E)u(x_*)\\ &= (1-\pi_A(E))u(z_A) + \pi_A(E)u(x_*).\end{aligned}$$

Therefore, $f_{P\setminus E} \sim_A x_* E z_A$. $\square$

Lemma 1.7.3. *Suppose $\mathcal{P} \neq \{\Omega, \emptyset\}$. Then $E \in \mathcal{E}$ if and only if $E \in \sigma(\mathcal{P})$.*

Proof. Suppose $E \notin \sigma(\mathcal{P})$. It needs to be shown that $E \notin \mathcal{E}$. There are two cases to consider.

Case 1: $E \subsetneq P$ for some $P \in \mathcal{P}$. Let $x^* \succ x_*$ be given and define $f_{P\setminus E} = x_* E(x^* P x_*)$. Let z be such that $f_{P\setminus E} \sim x_* E z$ as in the previous lemma, and let $\bar{z}$ be such that $f_P = x^* P x_* \sim \bar{z}$.

$$f_P = \begin{pmatrix} x^* & \text{if } \omega \in E \\ x^* & \text{if } \omega \in P\setminus E \\ x_* & \text{if } \omega \in \Omega\setminus P \end{pmatrix} \sim \bar{z} \quad \text{and} \quad f_{P\setminus E} = \begin{pmatrix} x_* & \text{if } \omega \in E \\ x^* & \text{if } \omega \in P\setminus E \\ x_* & \text{if } \omega \in \Omega\setminus P \end{pmatrix} \sim \begin{pmatrix} x_* & \text{if } \omega \in E \\ z & \text{if } \omega \in P\setminus E \\ z & \text{if } \omega \in \Omega\setminus P \end{pmatrix}$$

Since each state in Ω is non-null, $\pi(P) > \pi(P\setminus E)/(1-\pi(E))$. Hence, $u(x^*) > u(x_*)$ implies

$$u(\bar{z}) = \pi(P)u(x^*) + (1-\pi(P))u(x_*) > \frac{\pi(P\setminus E)}{1-\pi(E)}u(x^*) + \frac{1-\pi(P)}{1-\pi(E)}u(x_*) = u(z).$$

Therefore, by the representation, $\bar{z} \succ z$ and $\bar{z} \succ_{E^c} z$. Moreover, since π_{E^c} is the RML posterior of π, $\pi_{E^c}(P) = \pi(P)$. This implies that $f_P \sim_{E^c} \bar{z}$ as

$$U_{E^c}(f_P) = \pi_{E^c}(P)u(x^*) + (1-\pi_{E^c}(P))u(x_*) = u(\bar{z}).$$

In addition, since $\pi_{E^c}(E) = 0$, $f_{P\setminus E} \sim_{E^c} f_P \sim_{E^c} \bar{z}$ and $z \sim_{E^c} x_* E z$. Therefore, $\bar{z} \succ_{E^c} z$ implies $f_{P\setminus E} \succ_{E^c} x_* E z$. On the other hand, $f_{P\setminus E} \sim x_* E z$ and $f_{P\setminus E}$ and $x_* E z$ agree on E, violating dynamic consistency. Since E^c is not dynamically consistent, neither E nor E^c belongs to $\mathcal{E}$.

Case 2: There are $P, P' \in \mathcal{P}$ such that $P \cap E \neq \emptyset$, $P' \cap E \neq \emptyset$ and either $P \cap E^c \neq \emptyset$ or $P' \cap E^c \neq \emptyset$. Without loss of generality, assume that $P \cap E^c \neq \emptyset$. Let $A = E \cup P$. As before let $x^* \succ x_*$ be given. Consider a bet on $P \cap E$ given by $f_{P\cap E} = x_* P \setminus E(x^* P x_*)$. Let z_A be as in the previous lemma such that $x_* P \setminus E(x^* P x_*) \sim_A x_* P \setminus E z_A$, and let $\bar{z}_A \in X$ be such that $f_P = x^* P x_* \sim_A \bar{z}_A$. Then as in the previous case $\bar{z}_A \succ_A z_A$.

$$f_P = \begin{pmatrix} x^* & \text{if } \omega \in P \setminus E \\ x^* & \text{if } \omega \in P \cap E \\ x_* & \text{if } \omega \in \Omega \setminus P \end{pmatrix} \sim_A \bar{z}_A, \quad f_{P\cap E} = \begin{pmatrix} x_* & \text{if } \omega \in P \setminus E \\ x^* & \text{if } \omega \in P \cap E \\ x_* & \text{if } \omega \in \Omega \setminus P \end{pmatrix} \sim_A \begin{pmatrix} x_* & \text{if } \omega \in P \setminus E \\ z_A & \text{if } \omega \in P \cap E \\ z_A & \text{if } \omega \in \Omega \setminus P \end{pmatrix}$$

Since π_E is the RML posterior of π_A with the partition $\{A \cap P | P \in \mathcal{P}\}$ and $P'' \cap A \neq \emptyset$ implies $P'' \cap E \neq \emptyset$, it must be that $\pi_E(P) = \pi_A(P)$. Therefore, $f_{P\cap E} = x_* P \setminus E(x^* P x_*) \sim_E \bar{z}_A$. On the other hand, $x_* P \setminus E z_A \sim_E z_A$. Hence, $x_* P \setminus E(x^* P x_*) \sim_A x_* P \setminus E z_A$ but $x_* P \setminus E(x^* P x_*) \succ_E x_* P \setminus E z_A$, violating perfect dynamic consistency. This proves that $E \notin \mathcal{E}$.

Now suppose $E \in \sigma(\mathcal{P})$. It needs to be shown that $E \in \mathcal{E}$. Let $A \supseteq E$ be given. By Lemma 1.7.1, π_E is the RML posterior of π_A with the partition $\{A \cap P |\, P \in \mathcal{P}\}$. To prove that $E \in \mathcal{E}$, it suffices to show that for any $\omega \in E$,

$$\pi_E(\omega) = \frac{\pi_A(\omega)}{\pi_A(E)}.$$

Since $E \in \sigma(\mathcal{P})$, $\cup_{P \in \mathcal{P}:\, E \cap P \neq \emptyset} P = E$. By Lemma 1.2.1 and Lemma 1.7.1,

$$\pi_E(\omega) = \pi_A(\omega | P_\omega) \cdot \pi_A(P_\omega | \cup_{P \in \mathcal{P}:\, E \cap P \neq \emptyset} P) = \pi_A(\omega | P_\omega) \cdot \pi_A(P_\omega | E) = \frac{\pi_A(\omega)}{\pi_A(E)}$$

as desired. This concludes the proof of the lemma. □

Since $E \in \mathcal{E}$ if and only if $E \in \sigma(\mathcal{P})$, the necessity of Axiom 1.3.8 is obvious. To see the necessity of Axiom 1.3.9, let $A \in \mathcal{A}$ and $D \subsetneq A \cap P$ for some $P \in \mathcal{P}$. Since $\pi_{A \setminus D}$ is the RML posterior of π_A with the partition $\{A \cap P | P \in \mathcal{P}\}$, we have $\pi_{A \setminus D}(P) = \pi_A(P)$. Hence, Axiom

1.3.9 follows. Finally, to see the necessity of Axiom 1.3.10, let $D \subsetneq P$ for some $P \in \mathcal{P}$, $A \supseteq D$ and $\omega \in D$ be given. As π_D is the RML posterior of π_A,

$$\pi_D(\omega) = \pi_A(\omega|D) \cdot \pi_A(P|P) = \pi_A(\omega|D).$$

Hence, D satisfies perfect dynamic consistency.

Sufficiency

The first lemma is a standard result. For a proof, see Fishburn (1970) or Kreps (1988).

Lemma 1.7.4. *Suppose Axioms 1.3.1–1.3.5 are satisfied. Then for any $A \in \mathcal{A}$, there exist a subjective probability measure $\pi_A \in \Delta(\Omega)$ and a non-constant, affine utility function $u_A : X \to \mathbb{R}$ such that for any $f, g \in \mathcal{F}$,*

$$f \succcurlyeq_A g \Leftrightarrow \sum_{\omega \in \Omega} \pi_A(\omega) u_A(f(\omega)) \geq \sum_{\omega \in \Omega} \pi_A(\omega) u_A(g(\omega)).$$

By constant act preference invariance, $\succcurlyeq_A$ and $\succcurlyeq$ agree on all constant acts for all $A \in \mathcal{A} \setminus \emptyset$. Using the standard uniqueness result, for any $A \in \mathcal{A} \backslash \emptyset$, u_A is a positive affine transformation of u_Ω, which is denoted by u. Hence, it is without loss to let $u_A = u$ for all $A \in \mathcal{A} \setminus \emptyset$. Moreover, Axiom 1.3.4 and Axiom 1.3.6 imply that $\succcurlyeq_A$ is nontrivial, and hence π_A is unique for each $A \in \mathcal{A} \setminus \emptyset$ as in Anscombe and Aumann (1963). By consequentialism, for any $f \in \mathcal{F}$, $fAx^* \sim_A fAx_*$. Since by the representation $u(x^*) > u(x_*)$, we must have $\pi_A(A^c) = 0$. Moreover, by monotonicity, for any $\omega \in A$, $x^* \succ_A x_* \omega x^*$. The representation implies that $\pi_A(\omega) > 0$ for all $\omega \in A$. Hence, this establishes the following lemma.

Lemma 1.7.5. *Suppose Axioms 1.3.1–1.3.7 are satisfied. Then for any $A \in \mathcal{A}$, there exist a subjective probability measure $\pi_A \in \Delta(\Omega)$ and a non-constant, affine utility function $u : X \to \mathbb{R}$*

such that $u(X) = [u(x_*), u(x^*)]$, *and for any* $f, g \in \mathcal{F}$,

$$f \succcurlyeq_A g \Leftrightarrow \sum_{\omega \in \Omega} \pi_A(\omega) u(f(\omega)) \geq \sum_{\omega \in \Omega} \pi_A(\omega) u(g(\omega)).$$

Moreover, π_A *has full support on* A *and is unique for all* $A \in \mathcal{A} \backslash \emptyset$*, and* u *is unique up to a positive affine transformation.*

Let $\mathcal{E}$ be the collection of events which are perfectly dynamically consistent and whose complements are also perfectly dynamically consistent as in Definition 1.3.3. By definition, $\mathcal{E}$ is closed under complements. By Axiom 1.3.8, $\mathcal{E}$ is closed under intersections. Moreover, $\Omega \in \mathcal{E}$. Hence, $\mathcal{E}$ is an algebra over Ω. Let $\mathcal{P}_{\mathcal{E}}$ be the partitioning of the state space that generates $\mathcal{E}$, and $(\pi, \mathcal{P}_{\mathcal{E}})$ is the probability assessment. To establish the representation, it needs to be shown that each π_A is the RML posterior of π.

Lemma 1.7.6. *For any non-null* $A \in \mathcal{A}$*,* π_A *is the RML posterior of* π*.*

Proof. First, consider $A \in \mathcal{E}$. By definition, $f \succcurlyeq_A g \Leftrightarrow fAg \succcurlyeq g$. $fAg \succcurlyeq g$ is equivalent to

$$\sum_{\omega \in A} \pi(\omega) u(f(\omega)) \geq \sum_{\omega \in A} \pi(\omega) u(g(\omega)) \quad \Leftrightarrow \quad \sum_{\omega \in A} \frac{\pi(\omega)}{\pi(A)} u(f(\omega)) \geq \sum_{\omega \in A} \frac{\pi(\omega)}{\pi(A)} u(g(\omega)).$$

By the uniqueness of π_A in the representation, for all $\omega \in A$,

$$\pi_A(\omega) = \frac{\pi(\omega)}{\pi(A)} = \pi(\omega | A).$$

That is, π_A is the Bayesian posterior of π, which corresponds to the RML posterior as A is unambiguous.

Now consider $A \notin \mathcal{E}$. Let $\mathcal{P}_{\mathcal{E}} = \{P_1, \ldots, P_n\}$ and choose an index set $J \subseteq \{1, \ldots, n\}$ such that $P_j \cap A \neq \emptyset \Leftrightarrow j \in J$. Let $B = \cup_{j \in J} P_j$. Since $B \in \mathcal{E}$, by the first part of the lemma, π_B is the Bayesian posterior of π. Moreover, by Axiom 1.3.9, for any $j \in J$, $c_A(f_{P_j}) = c_{A \cup P_j}(f_{P_j})$ where $f_{P_j} = x^* P_j x_*$. Given the representation by Lemma 1.7.5, this is possible only if $\pi_A(P_j) =$

$\pi_{A\cup P_j}(P_j)$. Hence, iterative application Axiom 1.3.9 implies that $\pi_A(P_j) = \pi_B(P_j)$ for all $j \in J$. Therefore, for any $j, j' \in J$,

$$\frac{\pi_A(P_j)}{\pi_A(P_{j'})} = \frac{\pi_B(P_j)}{\pi_B(P_{j'})} = \frac{\pi(P_j)}{\pi(P_{j'})}.$$

Let $A \cap P_j$ for some $j \in J$ be given. By Axiom 1.3.10, $A \cap P_j$ is a perfectly dynamically consistent event. Hence, using the same reasoning as above, we get

$$\pi(\omega | A \cap P_j) = \pi_{A\cap P_j}(\omega) = \pi_A(\omega | P_j).$$

But then for any $\omega, \omega' \in A \cap P_j$,

$$\frac{\pi(\omega)}{\pi(\omega')} = \frac{\pi_{A\cap P_j}(\omega)}{\pi_{A\cap P_j}(\omega')} = \frac{\pi_A(\omega)}{\pi_A(\omega')}.$$

This together with the conclusion of the previous paragraph and Lemma 1.2.1 imply that π_A is the RML posterior of π. $\square$

Let $(\pi, \mathcal{P}, u)$ represent $\{\succcurlyeq_A\}_{A\in\mathcal{A}}$. To show that $\mathcal{P}$ is uniquely revealed as $\mathcal{P}_{\mathcal{E}}$, assume that $\mathcal{P}$ is not degenerate ($\mathcal{P} \neq \{\Omega, \emptyset\}$). By Lemma 1.7.3, $\mathcal{P}$ and $\mathcal{P}_{\mathcal{E}}$ are two partitions of the state space that generate the same algebra $\mathcal{E}$. But then $\mathcal{P} = \mathcal{P}_{\mathcal{E}}$. $\square$

Proof of Lemma 1.5.1

Let $(\pi, \mathcal{P}, u)$ be a representation of $\{\succcurlyeq_A\}_{A\in\mathcal{A}}$ given by Theorem 1.5.1. Observe that for any A,

$$\pi_A \in B(\mathbb{N}_{\mathcal{P},A}(\pi)) \Leftrightarrow \pi_A(A^c) = 0 \text{ and } \pi_A(P) = \frac{\pi(P)}{\sum_{P'\in\mathcal{P}:P'\cap A\neq\emptyset} \pi(P')}.$$

That is, all plausible posteriors agree on minimal unambiguous events. Moreover, the utility function defined by

$$U_A(f) = \sum_{P\in\mathcal{P}} \pi_A(P) \min_{\omega\in A\cap P} u(f(\omega)),$$

where $\pi_A \in B(\mathbb{N}_{\mathcal{P},A}(\pi))$, represents $\succcurlyeq_A$.

Let E be a dynamically consistent event. Suppose E is not perfectly dynamically consistent so that there exist $A \supseteq E$ and $f, g \in \mathcal{F}$ such that $fEg \succsim_A g$ and $g \succ_E f$. Let h and h' be given as below.

$$h = \begin{pmatrix} f(\omega) & \text{if } \omega \in E \\ g(\omega) & \text{if } \omega \in A \setminus E \\ x^* & \text{if } \omega \in A^c \end{pmatrix} \quad \text{and} \quad h' = \begin{pmatrix} g(\omega) & \text{if } \omega \in E \\ g(\omega) & \text{if } \omega \in A \setminus E \\ x^* & \text{if } \omega \in A^c \end{pmatrix}$$

Then, $h \succsim_A h'$, $h(\omega) = h'(\omega)$ for all $\omega \in E^c$ but $h' \succ_E h$.

Next, it is shown that $h \succsim_A h'$ implies $h \succsim h'$. Let $\mathcal{P} = \{P_1, \ldots, P_n\}$ and $J \subseteq \{1, \ldots, n\}$ be the index set such that $j \in J \Leftrightarrow P_j \cap E \neq \emptyset$. Then, by the representation, $h \succsim_A h'$ implies

$$\sum_{j \in J} \pi_A(P_j) \min_{\omega \in A \cap P_j} u(fEg(\omega)) \geq \sum_{j \in J} \pi_A(P_j) \min_{\omega \in A \cap P_j} u(g(\omega)).$$

Now for any $i \notin J$, $P_i \cap E = \emptyset$, and hence $\min_{\omega \in P_i} u(h) = \min_{\omega \in P_i} u(h')$. Moreover, since $h(\omega) = h'(\omega) = x^*$ for all $\omega \in A^c$, for any $j \in J$, $\min_{\omega \in A \cap P_j} u(h(\omega)) = \min_{\omega \in P_j} u(h(\omega))$ and $\min_{\omega \in A \cap P_j} u(h'(\omega)) = \min_{\omega \in P_j} u(h'(\omega))$. Finally, for any $j \in J$, $\pi(P_j) = c \cdot \pi_A(P_j)$ where $c = \sum_{P' \in \mathcal{P} : P' \cap A \neq \emptyset} \pi(P')$. Hence,

$$\begin{aligned} \sum_{j \in J} \pi_A(P_j) \min_{\omega \in A \cap P_j} u(fEg(\omega)) &\geq \sum_{j \in J} \pi_A(P_j) \min_{\omega \in A \cap P_j} u(g(\omega)) \Rightarrow \\ \sum_{j \in J} c \cdot \pi_A(P_j) \min_{\omega \in A \cap P_j} u(h(\omega)) &\geq \sum_{j \in J} c \cdot \pi_A(P_j) \min_{\omega \in A \cap P_j} u(h'(\omega)) \Rightarrow \\ \sum_{j \in J} \pi(P_j) \min_{\omega \in P_j} u(h(\omega)) &\geq \sum_{j \in J} \pi(P_j) \min_{\omega \in P_j} u(h'(\omega)) \Rightarrow \\ \sum_{P_i \in \mathcal{P}} \pi(P_i) \min_{\omega \in P_i} u(h(\omega)) &\geq \sum_{P_i \in \mathcal{P}} \pi(P_i) \min_{\omega \in P_i} u(h'(\omega)), \end{aligned}$$

which implies $h \succsim h'$. But then since $h(\omega) = h'(\omega)$ for all $\omega \in E^c$ and $h' \succ_E h$, this contradicts the original hypothesis that E is dynamically consistent. $\square$

Proof of Theorem 1.5.1

Necessity

The necessity of Axioms 1.3.1, 1.3.2, 1.3.4, 1.3.5, 1.3.6, and 1.3.7 is obvious. To prove the necessity of Axioms 1.3.8, 1.3.9, 1.5.1, and 1.5.2, it is shown that $\mathcal{E} = \sigma(\mathcal{P})$, where $\sigma(\mathcal{P})$ is the algebra generated by $\mathcal{P}$, as in the proof of Theorem 1.3.1.

Lemma 1.7.7. *Let $(\pi, \mathcal{P}, u)$ be a representation of $\{\succsim_A\}_{A \in \mathcal{A}}$ given by (1.5). Then $E \in \mathcal{E}$ if and only if $E \in \sigma(\mathcal{P})$.*

Proof. First, I show that if $E \notin \sigma(\mathcal{P})$, then $E \notin \mathcal{E}$. Notice that since $E \notin \sigma(\mathcal{P})$, there exists $P \in \mathcal{P}$ such that $P \cap E \neq \emptyset$ and $P \setminus E \neq \emptyset$. Let $f = x^* P \cap E x_*$. Then the representation implies that $f \sim x_*$ but $f \succ_E x_*$ even though $f(\omega) = g(\omega)$ for $\omega \in E^c$. $f \succ_E x_*$ holds because of the assumption that $\pi(P) > 0$ which implies $\pi_E(P) > 0$ for all $\pi_E \in B(\mathbb{N}_{\mathcal{P},E}(\pi))$.

Now suppose $E \in \sigma(\mathcal{P})$. To show that $E \in \mathcal{E}$, let $A \supseteq E$ be given. It needs to be shown that $fEg \succsim_A g$ if and only if $f \succsim_E g$. Let $\mathcal{P} = \{P_1, \ldots, P_n\}$. Since $E \in \sigma(\mathcal{P})$, there exists an index set $J \subseteq \{1, \ldots, n\}$ such that $E = \cup_{j \in J} P_j$. Notice that if $i \notin J$, then $\min_{\omega \in P_i} u(fEg(\omega)) = \min_{\omega \in P_i} u(g(\omega))$. Hence, $fEg \succsim_A g$ if and only if

$$\sum_{j \in J} \pi_A(P_j) \min_{\omega \in P_j} u(fEg(\omega)) \geq \sum_{j \in J} \pi_A(P_j) \min_{\omega \in P_j} u(g(\omega))$$

where π_A is an arbitrary member of $B(\mathbb{N}_{\mathcal{P},A}(\pi))$ as all plausible posteriors agree on minimal unambiguous events. On the other hand, it is easy to see that for any $\pi_E \in B(\mathbb{N}_{\mathcal{P},E}(\pi))$, $\pi_E(P_j) > 0$ if and only if $j \in J$, and $\pi_E(P_j) = c \cdot \pi_A(P_j)$ where $c = \frac{1}{\sum_{j \in J} \pi_A(P_j)}$. Hence, the above inequality holds if and only if

$$\sum_{j \in J} \pi_E(P_j) \min_{\omega \in P_j} u(fEg(\omega)) \geq \sum_{j \in J} \pi_E(P_j) \min_{\omega \in P_j} u(g(\omega))$$

which is true if and only if $f \succsim_E g$. $\square$

Since $\mathcal{E} = \sigma(\mathcal{P})$, necessity of Axiom 1.3.8 is obvious. The necessity of Axiom 1.3.9 is the same as in Theorem 1.3.1. To see the necessity of Axiom 1.5.1, let S be an alternative state space such that the mapping $\Gamma : \Omega \to S$, where Γ satisfies $\Gamma(\omega) = \Gamma(\omega')$ for all $\omega, \omega' \in P$ and $\Gamma(\omega) \neq \Gamma(\omega')$ whenever $\omega \in P$ and $\omega' \in P'$ for distinct P and P', is a surjection. Now for each $A \in \mathcal{A}$, define a probability measure on S by $\pi_A \circ \Gamma^{-1}$ where $\pi_A \in B(\mathbb{N}_{\mathcal{P},A}(\pi))$. Let $\hat{\mathcal{F}}$ be the set of all acts X^S. The set $\hat{\mathcal{F}}$ is isomorphic $\mathcal{F}^{ua}$. Since $(\pi_A \circ \Gamma^{-1}, u)$ is an SEU representation of $\succcurlyeq_A$ restricted to $\hat{\mathcal{F}}$, Axiom 1.5.1 follows. Finally, Axiom 1.5.2 is necessary as in each partition only the minimal payoff matters.

Sufficiency

Axiom 1.3.8 implies that $\mathcal{E}$ is an algebra and $\mathcal{P}_{\mathcal{E}}$ is a partitioning of the state space. It is easy to see that Axioms 1.3.1, 1.3.2, 1.3.4, 1.3.5, 1.3.6, and 1.5.1 imply the following lemma.

Lemma 1.7.8. *For any $A \in \mathcal{A}$, there exist a subjective probability measure π_A on $\sigma(\mathcal{P}_{\mathcal{E}})$ and a non-constant, affine utility function $u : X \to \mathbb{R}$ such that for any $f, g \in \mathcal{F}^{ua}$,*

$$f \succcurlyeq_A g \Leftrightarrow \sum_{P \in \mathcal{P}} \pi_A(P) u(f(P)) \geq \sum_{P \in \mathcal{P}} \pi_A(P) u(g(P)).$$

By consequentialism, $\pi_A(P) = 0$ whenever $A \cap P = \emptyset$. Extend π_A to $\sigma(\mathcal{P}_{\mathcal{E}} \cup A)$ — the algebra generated by sets of the form $A \setminus P$, $A \cap P$, and $P \setminus A$, by letting $\pi_A(A \cap P) = \pi_A(P)$ whenever $A \cap P \neq \emptyset$.

Lemma 1.7.9. *For any $A \in \mathcal{A}$ and $f \in \mathcal{F}$, there exists $f^{ua} \in \mathcal{F}^{ua}$ such that $f \sim_A f^{ua}$.*

Proof. Let $P \in \mathcal{P}_{\mathcal{E}}$ and $\omega^* = \arg\min_{\omega \in A \cap P} u(f(\omega))$. By Axioms 1.3.7 and 1.5.2, $x^* P \setminus \{\omega^*\} f \sim_A f(\omega^*) P f$. On the other hand, by monotonicity, $x^* P \setminus \{\omega^*\} f \succcurlyeq_A f \succcurlyeq_A f(\omega^*) P f$. Hence, $f \sim_A f(\omega^*) P f$. Now let f^{ua} denote an act that assigns the worst prize of f in $A \cap P$ to P for all $P \in \mathcal{P}_{\mathcal{E}}$ with $A \cap P \neq \emptyset$ and is constant on P' with $A \cap P' = \emptyset$. This act belongs to $\mathcal{F}^{ua}$, and $f \sim_A f^{ua}$ by iterative application of the previous argument and Axiom 1.3.7. $\square$

Now for any $f \in \mathcal{F}$, let

$$U_A(f) = \sum_{P \in \mathcal{P}} \pi_A(A \cap P) \min_{\omega \in A \cap P} u(f(\omega)).$$

Notice that for f^{ua} defined as in Lemma 1.7.9, $U_A(f) = U_A(f^{ua})$. We already know that U_A represents $\succsim_A$ on $\mathcal{F}^{ua}$. Hence, $f \succsim_A g$ if and only if $f^{ua} \succsim_A g^{ua}$ if and only if $U_A(f^{ua}) \geq U_A(g^{ua})$ if and only if $U_A(f) \geq U_A(g)$.

The only thing left to prove is that $\pi_A \in B(\mathbb{N}_{\mathcal{P},A}(\pi))$. This is implied by Axiom 1.3.9. The proof is identical to the first part of Lemma 1.7.6. Finally, the uniqueness result for u is standard. The uniqueness of $\mathcal{P}$ is a consequence of Lemma 1.7.7, and $\mathbb{N}_{\mathcal{P},A}(\pi)$ is uniquely defined given π on $\mathcal{P}$. $\square$

Proof of Proposition 1.2.1

Notice that $\pi'_A \in B(\mathbb{N}_{\mathcal{P},A}(\pi))$ if and only if for all $P \in \mathcal{P}$ with $A \cap P \neq \emptyset$,

$$\sum_{\omega \in A \cap P} \pi'_A(\omega) = \frac{\pi(P)}{\sum_{P' \in \mathcal{P}:\, A \cap P' \neq \emptyset} \pi(P')}. \tag{A.1}$$

The objective is to minimize Kullback-Leibler divergence $D_{\mathrm{KL}}(\pi(\cdot|A) \,||\, \pi'_A)$ subject to these constraints for each $P \in \mathcal{P}$ with $A \cap P \neq \emptyset$. The Lagrangian for the minimization problem is

$$\begin{aligned}\mathcal{L}\big(\{\pi'_A(\omega)\}_{\omega \in A}, \{\lambda_P\}_{P \in \mathcal{P}: A \cap P \neq \emptyset}\big) = &-\sum_{\omega \in A} \pi(\omega|A) \ln\Big(\frac{\pi'_A(\omega)}{\pi(\omega|A)}\Big) \\ &+ \sum_{P \in \mathcal{P}: A \cap P \neq \emptyset} \lambda_P \Big(\sum_{\omega \in A \cap P} \pi'_A(\omega) - \frac{\pi(P)}{\sum_{P' \in \mathcal{P}:\, A \cap P' \neq \emptyset} \pi(P')}\Big).\end{aligned}$$

The first order conditions imply that for any $P \in \mathcal{P}$ with $A \cap P \neq \emptyset$ and any $\omega, \omega' \in A \cap P$,

$$\frac{\pi(\omega|A)}{\pi'_A(\omega)} = \lambda_P = \frac{\pi(\omega'|A)}{\pi'_A(\omega')}, \quad \text{and hence} \quad \frac{\pi'_A(\omega)}{\pi'_A(\omega')} = \frac{\pi(\omega|A)}{\pi(\omega'|A)} = \frac{\pi(\omega)}{\pi(\omega')}. \tag{A.2}$$

Since the objective function is strictly convex, (A.1) and (A.2) characterize the solution to the minimization problem. □

CHAPTER 2

Stochastic Attention and Reference Dependent Choice

2.1 Introduction

Starting with the work of Kahneman and Tversky (1979) the idea that choices are reference dependent has played a significant role in the economics literature. Influenced by growing evidence on reference dependent choice, scholars have developed theories of choice behavior which can explain a wide variety of empirical findings (for example, see Tversky and Kahneman, 1991; Munro and Sugden, 2003; Sugden, 2003; Sagi, 2006; Masatlioglu and Ok, 2005, 2014; Apesteguia and Ballester, 2009). However, most existing models only address deterministic choice behavior. Since observed choices frequently exhibit randomness (Tversky, 1969; Agranov and Ortoleva, 2017), it is important to understand the foundations for reference dependent random choice behavior.

In this paper, I study reference dependent random choice. Randomness in choice is modeled through the attention channel. I assume that the Decision Maker (DM) pays attention to a *subset* of available alternatives, which is also called the DM's *consideration set*. Moreover, the set of alternatives the DM pays attention to is directly influenced by her reference point. The idea that consumers do not pay attention to all available alternatives has a long tradition in the marketing literature (Wright and Barbour, 1977; Hauser and Wernerfelt, 1990). In the economics literature, this has also been recognized as a fruitful avenue of research, and scholars have developed models which allow for revealed preference analysis in the presence of limited attention (e.g., Masatlioglu et al., 2012; Manzini and Mariotti, 2014; Aguiar, 2015, 2017; Brady and Rehbeck, 2016; Lleras et al., 2017; Cattaneo et al., 2019). The goal of this paper is to understand the scope of existing reference dependent random attention models in explaining observed choice behavior.

The main focus of this paper will be three reference dependent random attention models: fixed independent consideration (Manzini and Mariotti, 2014), logit consideration (Brady and Rehbeck, 2016), and fixed correlated consideration (Aguiar, 2017). In fixed independent consideration, each alternative has a fixed reference dependent probability of being considered that does not depend on the menu. Moreover, the fact that one alternative is considered has no bearing on the consideration probability of other alternatives. In fixed correlated consideration, each alternative still has a fixed reference dependent probability of being considered, but correlations between alternatives are allowed. In logit consideration, the consideration probability of each alternative depends on the choice set and correlations between alternatives are allowed, but it is assumed that the relative weights of two consideration sets do not depend on choice sets.

To understand the behavioral implications of these three models, I start with a fairly general reference dependent random attention model (RDRAM) which includes all of these models as special cases. In the general model, it is assumed that the DM has a fixed reference independent strict preference over alternatives and reference dependent random attention. The only structure imposed on the attention rule is that (i) the reference point is always considered, (ii) each subset of a given menu that includes the reference alternative has a positive probability of being the DM's consideration set. While this model is fairly general, it has nontrivial behavioral implications. The first implication is that the reference alternative always has a positive probability of being chosen. The second implication is that each menu has a unique dominant alternative — the alternative that is never abandoned when it is the reference point. The last property requires consistency of dominant alternatives across different choice sets.

In Section 2.3, I show that the two key properties of fixed independent consideration are *irrelevance of dominated alternatives* (IDA) and *ratio independence of dominant alternatives* (RIDA). Irrelevance of dominated alternatives says that if x is chosen with positive probability when y is the reference point, then given any choice problem (S, r), removing y from S cannot affect the probability that x is chosen. In other words, the removal of dominated alternatives from a menu cannot affect the choice probability of dominant alternatives. Ratio independence of dominant alternatives says that if x is never abandoned for any other alternative in a given menu when it is the reference point, then removing x from the menu cannot affect the the relative choice probability of any other two alternatives. Hence, when the dominant alternative in the menu is removed, the probability that it is chosen is distributed among the remaining alternatives in proportion to their original choice probabilities.

In Section 2.2, I show that a reference dependent random attention rule has a fixed independent consideration representation if and only if it has logit and fixed correlated consideration representations. This is one of the key results of this paper. Even though the result is stated in terms of the underlying attention rule, the equivalent result also holds for reference dependent random choice rules. In particular, the key property of the logit consideration model is ratio independence of dominant alternatives, whereas the key property of the fixed correlated consideration model is irrelevance of dominated alternatives. Hence, a suitable relaxation of one of the axioms underlying fixed independent consideration model results in logit or fixed correlated consideration model.

I also discuss recoverability of attention parameters from observed choices for logit and fixed correlated consideration models. If the reference alternative is the worst alternative among all alternatives, then identification of all attention parameters can be achieved in both models (also see Brady and Rehbeck, 2016; Aguiar, 2017). In the general case, identification of attention parameters (up to normalization) is still possible in the logit model. However, the degree of freedom is higher in fixed correlated consideration model when the reference alternative is not the worst alternative. The intuition for this is as follows. As mentioned earlier, fixed correlated attention model satisfies irrelevance of dominated alternatives property. Consider a choice problem (S, r) and suppose new alternatives which are dominated by the reference alternative r are introduced. Irrelevance of dominated alternatives requires that this will not change the choice probability of any alternative. Hence, observed choices in the original and new choice problems will be exactly the same and this additional data cannot be used to recover attention parameters.

This paper is contributes to the literature on reference dependent preferences and limited (random) attention. As discussed earlier, most reference dependent choice models are deterministic. In contrast, this paper completely focuses on random choice data. The most closely related papers on reference dependent preferences are Masatlioglu and Ok (2005, 2014). They consider a model where the decision maker maximizes her reference independent preference within a subset of all available alternatives. Their model is a limit case fixed correlated consideration where an alternative is considered either with zero or one probability. Dean, Kıbrıs, and Masatlioglu (2017) provide an extension of Masatlioglu and Ok (2014) where attention has both reference dependent and reference independent component. Their model is independent from all the models discussed in this paper.

The only paper that I am aware of which discusses reference dependent random choice is Kovach (2016). He provides axiomatic foundations for fixed independent consideration. His axiomatic structure looks quite different from the one in this paper, however the equivalence is not difficult to establish. He provides an in depth analysis of the fixed independent consideration model and its relationship with random utility models. In contrast, the focus in this paper is to provide a framework within which many other models of random attention are studied and to investigate the relationship between these models.

This paper also contributes to the literature on limited (random) attention. Fixed independent consideration with standard random choice data was first axiomatized by Manzini and Mariotti (2014) (see also Horan, 2013). In a similar setup, Brady and Rehbeck (2016) and Aguiar (2017) provide characterization results for logit and fixed correlated consideration models, respectively. This paper contributes to this literature by providing a precise relationship between these models and showing that the intersection of fixed correlated and logit consideration is exactly fixed independent consideration. Another contribution of this paper is to extend these random attention models to reference dependent choice framework and to illustrate to what degree these models can be used to explain empirical findings on reference dependence.

2.2 General Model

Throughout this paper X denotes the grand finite set of alternatives. $\mathcal{X}$ is the set of all nonempty subsets of X. A choice problem is (S, r) where $S \in \mathcal{X}$ is a menu of alternatives and $r \in S$ is a reference point. Reference dependent random choice is a collection $\{p_r\}_{r \in X}$, $p_r : X \times \mathcal{X} \to [0, 1]$ such that for any (S, r), $p_r(x, S) \geq 0$ for each $x \in S$ and $\sum_{x \in S} p_r(x, S) = 1$.

For expositional clarity, I will restrict attention to reference dependent random choices which satisfy the following three properties: Nontrivial Reference Effect, Dominant Alternative, and Consistency of Dominance. The first property says that the reference alternative is always chosen with positive probability. The second property states that in any choice set there exists unique dominant alternative, i.e., an alternative which is never abandoned for other alternatives when it is the reference point. The last property requires that if x is revealed to be the dominant alternative in some choice set that contains y, then in any choice set that includes x and y, y must be chosen with zero probability when x is the reference point. In addition, x must be chosen with positive probability when y is the reference point. These properties are listed below.

- *Nontrivial Reference Effect:* For any choice problem (S, r), $p_r(r, S) > 0$.
- *Dominant Alternative:* For any $S \in \mathcal{X}$, there exists unique $x^* \in S$ such that $p_{x^*}(x^*, S) = 1$.
- *Consistency of Dominance:* If $p_x(x, S) = 1$ for some $S \in \mathcal{X}$, then for any $y \in S \setminus x$ and $T \supseteq \{x, y\}$, $p_x(y, T) = 0$ and $p_y(x, T) > 0$.

I will now introduce a very general model the behavioral content of which is exactly these three properties. The aim is to provide a framework that is common to all reference dependent random attention models considered in this paper so that the distinguishing features of each of the special cases are clearer. In the general model each reference alternative has an associated random attention rule $\mu_r : \mathcal{X} \times \mathcal{X} \to [0, 1]$ such that $\mu_r(D, S) > 0$ if and only if $r \in D \subseteq S$ and $\sum_{D:\, r \in D \subseteq S} \mu_r(D, S) = 1$. Here, $\mu_r(D, S)$ reflects the probability that the DM has a consideration set $D \subseteq S$ when the choice problem (S, r). It is assumed that each $D \subseteq S$ containing the reference point has a strictly positive probability of being the consideration set. Given a choice problem (S, r), the decision maker first draws a consideration set from the distribution $\mu_r(\cdot|S)$ and maximizes her preference, which is a linear order, in the realized consideration set. A binary relation $\succ$ on X is a linear order if it is (i) total: for any $x \neq y$ in X, either $x \succ y$ or $y \succ x$, (ii) asymmetric: for any $x \neq y$ in X, $x \succ y$ implies $y \not\succ x$, and (iii) transitive: for any $x, y, z \in X$, $x \succ y \succ z$ implies $x \succ z$.

Definition 2.2.1. *Reference dependent random choice p_r has a random attention representation (alternatively, p_r is an RDRAM) if there exist a linear order $\succ$ on X and a random attention rule μ_r such that for any choice problem (S, r) and for any $x \in S$,*

$$p_r(x, S) = \sum_{D \subseteq S:\, x = \text{argmax}(\succ, D)} \mu_r(D, S). \tag{2.1}$$

The next theorem states that Nontrivial Reference Effect, Dominant Alternative and Consistency of Dominance are necessary and sufficient for any reference dependent stochastic choice p_r to have a random attention representation.

Theorem 2.2.1. *Stochastic choice p_r is an RDRAM if and only if it satisfies Nontrivial Reference Effect, Dominant Alternative and Consistency of Dominance.*

From now on I will restrict attention to reference dependent random choices which have a random attention representation. I will consider random attention models which impose more structure on μ_r and explore the properties of each of these models.

Suppose p_r has a random attention representation. We say that p_r has a

1. Fixed Independent Consideration (FIC) representation if there exists $\gamma_r : X \to [0, 1]$ such that $\gamma_r(r) = 1$, $\gamma_r(x) \in (0, 1)$ for each $x \in X \setminus r$, and

$$\mu_r(D, S) = \prod_{x \in D} \gamma_r(x) \prod_{y \in S \setminus D} (1 - \gamma_r(y)), \tag{2.2}$$

2. Logit Consideration (LC) representation if there exists $\pi_r : \mathcal{X} \to [0,1]$ such that $\pi_r(D) > 0$ whenever $r \in D \subseteq X$, $\sum_{D\colon r\in D\subseteq X} \pi_r(D) = 1$, and

$$\mu_r(D,S) = \frac{\pi_r(D)}{\sum_{D'\colon r\in D'\subseteq S} \pi_r(D')}, \tag{2.3}$$

3. Fixed Correlated Consideration (FCC) representation if there exists $\pi_r : \mathcal{X} \to [0,1]$ such that $\pi_r(D) > 0$ whenever $r \in D \subseteq X$, $\sum_{D\colon r\in D\subseteq X} \pi_r(D) = 1$, and

$$\mu_r(D,S) = \sum_{D'\colon D'\cap S = D} \pi_r(D'), \tag{2.4}$$

2.2.1 Relationship Between Models

In this subsection, I show that the intersection of logit and fixed correlation consideration models is exactly fixed independent consideration. This is formally stated in the next proposition.

Proposition 2.2.1. *A random attention rule μ_r has a fixed independent consideration representation if and only if it has logit and fixed correlated consideration representations.*

Proof. First, suppose that μ_r has a FIC representation, i.e., equation (2.2) is satisfied for some γ_r. Let $\pi_r(D) = \prod_{x\in D}\gamma_r(x)\prod_{y\in X\setminus D}(1-\gamma_r(y))$. Then,

$$\begin{aligned}
\frac{\pi_r(D)}{\sum_{D'\colon r\in D'\subseteq S}\pi_r(D')} &= \frac{\prod_{x\in D}\gamma_r(x)\prod_{y\in X\setminus D}(1-\gamma_r(y))}{\sum_{D'\colon r\in D'\subseteq S}\prod_{x\in D'}\gamma_r(x)\prod_{y\in X\setminus D'}(1-\gamma_r(y))} \\
&= \frac{\prod_{x\in D}\gamma_r(x)\prod_{y\in S\setminus D}(1-\gamma_r(y))}{\sum_{D'\colon r\in D'\subseteq S}\prod_{x\in D'}\gamma_r(x)\prod_{y\in S\setminus D'}(1-\gamma_r(y))} \\
&= \prod_{x\in D}\gamma_r(x)\prod_{y\in S\setminus D}(1-\gamma_r(y))
\end{aligned}$$

where the last equality follows from the fact that the denominator is 1. Hence, fixed independent consideration is a special case of logit attention. In addition,

$$\begin{aligned}
\sum_{D':D'\cap S=D}\pi_r(D') &= \sum_{D':D'\cap S=D}\prod_{x\in D'}\gamma_r(x)\prod_{y\in X\setminus D'}(1-\gamma_r(y))\\
&=\prod_{x\in D}\gamma_r(x)\prod_{y\in S\setminus D}(1-\gamma_r(y))\Big[\sum_{D':D'\cap S=D}\prod_{x\in D'\setminus D}\gamma_r(x)\prod_{y\in(X\setminus S)\setminus D'}(1-\gamma_r(y))\Big]\\
&=\prod_{x\in D}\gamma_r(x)\prod_{y\in S\setminus D}(1-\gamma_r(y))
\end{aligned}$$

where the last equality follows from the observation that the term inside big parentheses has to be equal to 1. To see this, multiply this term by $\gamma_r(r)$ to get that the term inside the big parentheses is equal to $\sum_{r\in D''\subseteq(X\setminus S)\cup r}\mu_r(D'',(X\setminus S)\cup r)=1$. Hence, fixed independent consideration is also a special case of fixed correlated consideration.

Now suppose μ_r has both logit attention and fixed correlated attention representations. That is there exist π_r satisfying equation (2.3) and π'_r satisfying equation (2.4). Let $\gamma_r(x) = \mu_r(\{r,x\},\{r,x\})$. It needs to be shown that γ_r satisfies equation (2.2). This is trivially satisfied when S is binary. Suppose the claim holds whenever S has cardinality less than k and let S with $|S|=k$ be given. First, suppose $D\subsetneq S$ and pick $x\in S\setminus D$. Since μ_r has a logit attention representation,

$$\mu_r(D,S)=\mu_r(D,S\setminus x)\sum_{D':\,r\in D'\subseteq S\setminus x}\mu_r(D',S).$$

By induction hypothesis,

$$\mu_r(D,S\setminus x)=\prod_{y\in D}\gamma_r(y)\prod_{z\in S\setminus(D\cup x)}(1-\gamma_r(z)).$$

Furthermore, as μ_r has a fixed correlated representation,

$$\sum_{D':\,r\in D'\subseteq S\setminus x}\mu_r(D',S)=\mu_r(\{r\},\{r,x\})=(1-\gamma_r(x)).$$

Hence, it follows that

$$\mu_r(D,S) = \prod_{y \in D} \gamma_r(y) \prod_{z \in S \setminus D} (1 - \gamma_r(z))$$

whenever $D \subsetneq S$. By using the fact that $\mu_r(S,S) = 1 - \sum_{D:\, r \in D \subsetneq S} \mu_r(D,S)$, the claim follows for all $D \subseteq S$. This concludes the proof of the proposition. □

2.3 Fixed Independent Consideration

In this section, I introduce two axioms which are necessary and sufficient for a fixed independent consideration representation. The first axiom says that if x dominates y in a binary comparison, then removing y from any choice set cannot affect the probability that x is chosen. This property is called Irrelevance of Dominated Alternatives (IDA).

Axiom 2.3.1. *(Irrelevance of Dominated Alternatives - IDA) If $p_x(x, \{x,y\}) = 1$, then for any S,*

$$p_r(x,S) = p_r(x, S \setminus y).$$

Since there are fewer alternatives in $S \setminus y$ compared to S, the property says that when y is removed from the choice set S the probability that it is chosen in S will be distributed among alternatives which are dominated by y. One implication of this property is that given a choice set if we add alternatives which are inferior to the reference alternative, then the probability that the reference alternative is chosen remains unaffected. In fact, the probability that any alternative is chosen should remain unaffected. The experimental evidence from Samuelson and Zeckhauser (1988) clearly violates this property. They compare two choice sets one in which there is only one alternative apart from the reference alternative and another choice set in which there are four alternatives two of which are dominated by the reference alternative. They show that in the latter choice set the reference alternative is more likely to be chosen.

To see why this axiom holds, notice that fixed independent consideration model implies that

$$\mu_r(D, S \setminus y) = \mu_r(D \cup y, S) + \mu_r(D, S).$$

That is, the probability that D is the consideration set in the choice set $S \setminus y$ is equal to probability that either D or $D \cup y$ is the consideration set in S. If x dominates y, then x is the best alternative in $D \subseteq S \setminus y$ if and only if it is best in $D \cup y$. This implies that removing y from the menu cannot affect the probability that x will be chosen.

The next property relates the relative choice probabilities of two alternatives when the best alternative in the choice set is removed from it. It says that if x is the dominant alternative in S, then for any two alternatives $y, z \in S \setminus x$ which are chosen with positive probability, the relative probability that y is chosen as opposed to z stays the same when x is removed from the menu. This is a weak version of Luce's Independence of Irrelevant Alternatives axiom. Recall that Luce's IIA states that this ratio stays the same when we compare arbitrary choice sets S and T which include y and z. In fixed independent consideration model this is no longer true in general. Only dominant alternatives can be considered irrelevant for relative probabilities.

Axiom 2.3.2. *(Ratio Independence of Dominant Alternatives - RIDA). If $p_x(x, S) = 1$, then*

$$\frac{p_r(y, S)}{p_r(z, S)} = \frac{p_r(y, S \setminus x)}{p_r(z, S \setminus x)}$$

for any $y, z \in S \setminus x$ with $p_r(y, S) > 0$ and $p_r(z, S) > 0$.

To understand the property better, notice that RIDA imposes a structure on how the probability of x being chosen is distributed among other alternatives when it is removed from the menu. In particular, it states that the distribution must be proportional to the original probability that alternatives were chosen before x was removed from the menu. To illustrate the type of choice behavior that RIDA rules out, consider a choice problem $(\{r, x, y, z\}, r)$ where x is an alternative very similar to y, which it dominates in terms of preference, and z is very dissimilar to x. The similarity effect hypothesis (Tversky, 1972) states that in this type of choice situations y is disproportionately negatively affected by the addition of x to the menu. Hence, the relative probability that y is chosen as opposed to z is likely to be higher when x is not in the menu. However, fixed independent consideration rules out this type of effects.

To see why this property holds in fixed independent consideration, first notice that as x is the dominant alternative in S, removing it from the menu does not alter which consideration sets have y or z as the best alternative. Moreover, fixed independent consideration implies that the weight of each consideration set which has y or z as the best alternative is boosted by the same constant when x is removed from the menu. Thus the relative probability that y is chosen as opposed to z stays the same.

The next theorem provides a characterization result for fixed independent consideration.

Theorem 2.3.1. *Suppose p_r is an RDRAM. Then p_r has a fixed independent consideration representation if and only if it satisfies Irrelevance of Dominated Alternatives and Ratio Independence of Dominant Alternatives.*

The proof of the theorem involves three steps. First, an immediate implication of RIDA is that for any (S, r), the ratio $p_r(y, S)/p_r(y, S \setminus x)$ is constant for all $y \in S$ which are chosen with positive probability given that x is the dominant alternative in S. IDA is used to show that this constant is exactly equal to $p_r(r, \{x, r\})$. $\gamma_r(x)$ is defined to be the maximum of $p_r(x, \{x, r\})$ and $p_x(r, \{x, r\})$. The proof then follows by showing the representation for binary choice sets and using induction together with the previous conclusion to extend the representation to all choice sets. Notice that in the proof of the theorem $\gamma_r(x)$ is defined to be symmetric in the sense that $\gamma_r(x) = \gamma_x(r)$ for all $r, x \in X$. That is because $\gamma_r(x)$ is not identifiable from choice data whenever $r \succ x$. Hence, if p_r has a fixed independent consideration representation, then it is without loss to assume that γ_r is symmetric.

2.4 Logit Consideration

In this section, I explore the logit consideration model. As discussed in Section 2.2, logit consideration is a generalization of fixed independent consideration. The next two examples illustrate that (i) Irrelevance of Dominated Alternatives is no longer satisfied in logit consideration, and (ii) Ratio Independence of Dominant Alternatives by itself is not sufficient for logit consideration representation. The intuition for why RIDA holds in the logit model is exactly the same as in the fixed independent consideration model.

Example 2.4.1 (IDA Violation)**.** *Let* $x \succ y \succ z$ *and* $\pi_y(\{y\}) = \pi_y(\{x, y\}) = \pi_y(\{x, y, z\}) = 0.2$*,* $\pi_y(\{y, z\}) = 0.4$*. In the logit consideration model the implied choice probabilities when* y *is the reference alternative are as below.*

$p_y(\cdot, S)$	$S = \{x, y, z\}$	$\{x, y\}$	$\{y, z\}$
x	0.4	0.5	
y	0.6	0.5	1
z	0		0

IDA is violated as $p_y(y, \{x, y, z\}) \neq p_y(y, \{x, y\})$*.*

Example 2.4.2 (Insufficiency of RIDA)**.** *Let* $\{x, y, z\}$ *be the grand set of alternatives and assume that* p_r *is an RDRAM with* x *as the dominant alternative. When* z *is the reference alternative the choice data is as below.*

$p_z(\cdot, S)$	$S = \{x, y, z\}$	$\{x, z\}$	$\{y, z\}$
x	0.1	0.5	
y	0.45		0.5
z	0.45	0.5	0.5

Since p_r *is an RDRAM,* z *must be the worst alternative in* $\{x, y, z\}$*. The data given above satisfies RIDA as* $\frac{p_z(y,\{x,y,z\})}{p_z(z,\{x,y,z\})} = 1 = \frac{p_z(y,\{y,z\})}{p_z(z,\{y,z\})}$*. Now suppose that* p_z *has a logit attention represen-*

tation. Then $\frac{p_z(x,\{x,z\})}{p_z(z,\{x,z\})} = 1 = \frac{\pi_z(\{x,z\})}{\pi_z(\{z\})}$ *and* $\frac{p_z(x,\{x,y,z\})}{p_z(z,\{x,y,z\})} = \frac{2}{9} = \frac{\pi_z(\{x,z\})+\pi_z(\{x,y,z\})}{\pi_z(\{z\})}$*. Obviously, there cannot exist* π_z *which is a probability and satisfies these two equations.*

Example 2.4.2 illustrates why RIDA is not sufficient. Consider the following π_z.

D	$\{x,y,z\}$	$\{x,z\}$	$\{y,z\}$	$\{z\}$
$\pi_z(D)$	$-7/9$	1	1	1

Assuming that logit consideration holds, this π_z is the unique (up to normalization) way to represent the given choice behavior. However, this π_z is not a probability distribution. Hence, an extra assumption on p_r is needed so that π_r derived from observed choices is a well defined probability.

2.4.1 Identification of Consideration Probabilities

Before introducing the final axiom I discuss identification of consideration probabilities from choice data. The following theorem will be useful for this purpose.

Theorem 2.4.1. *(Möbius Inversion (Shafer, 1976)) If* Θ *is a finite set and* f *and* g *are functions on* 2^{Θ}*, then*

$$f(A) = \sum_{B \subseteq A} g(B)$$

for all $A \subseteq \Theta$ *if and only if*

$$g(A) = \sum_{B \subseteq A} (-1)^{|A \setminus B|} f(B)$$

for all $A \subseteq \Theta$*.*

First, suppose r is the worst alternative in X. Notice that if p_r has a logit consideration representation, then for any choice set $S \subseteq X$, it has to be the case that

$$\frac{p_r(r,X)}{p_r(r,S)} = \sum_{D:\, r \in D \subseteq S} \pi_r(D).$$

Then, Möbius Inversion implies

$$\pi_r(S) = \sum_{D\colon r\in D\subseteq S} (-1)^{|S\setminus D|} \frac{p_r(r,X)}{p_r(r,D)}.$$

Hence, unique identification of π_r from choice data is possible when r is the worst alternative. This is not true if r is not the worst alternative. For example, if r is the best alternative, then any choice of π_r works as $p_r(r,S) = 1$ for all S.

The general methodology for constructing π_r when r is arbitrary is described in the Appendix. Here I illustrate the case when r is the second worst alternative in the grand set of alternatives. Let s denote the alternative worse than r and $\mathcal{D}_r = \{\{r\},\{r,s\}\}$. P_r represents the set of alternatives preferred to r, i.e., $P_r = X \setminus \{r,s\}$. Given any π_r, we can define two probability distributions π_r^r and $\pi_r^{\{r,s\}}$ derived from π_r. π_r^r assigns positive probability only to subsets of X which do not include s. For any S such that $r \in S \subseteq X \setminus s$, define

$$\pi_r^r(S) = \frac{\pi_r(S)}{\sum_{D\colon r\in D\subseteq X\setminus s} \pi_r(D)}.$$

On the other hand, $\pi_r^{\{r,s\}}$ assigns positive probability only to subsets of X that include both r and s. For any S such that $\{r,s\} \subseteq S \subseteq X$, let

$$\pi_r^{\{r,s\}}(S) = \pi_r(S\setminus s) + \pi_r(S).$$

Now consider the set $X\setminus s$. If p_r has a logit attention representation, then the relative probability that the reference alternative is chosen in $X \setminus s$ as opposed to any other subset of $X \setminus s$ is given by

$$\frac{p_r(r,X\setminus s)}{p_r(r,S)} = \frac{\sum_{D\colon r\in D\subseteq S} \pi_r(D)}{\sum_{D'\colon r\in D'\subseteq X\setminus s} \pi_r(D')} = \sum_{D\colon r\in D\subseteq S} \pi_r^r(D).$$

The relative probability that the reference alternative is chosen in X as opposed any other set S

containing $\{r, s\}$ is given by

$$\frac{p_r(r, X)}{p_r(r, S)} = \sum_{D\colon \{r,s\} \subseteq D \subseteq S} (\pi_r(D \setminus s) + \pi_r(D)) = \sum_{D\colon \{r,s\} \subseteq D \subseteq S} \pi_r^{\{r,s\}}(D).$$

We can use Möbius Inversion to back out both π_r^r and $\pi_r^{\{r,s\}}$ from observed choices. For any S with $r \in S \subseteq X \setminus s$,

$$\pi_r^r(S) = \sum_{D\colon r \in D \subseteq S} (-1)^{|S \setminus D|} \frac{p_r(r, X \setminus s)}{p_r(r, D)},$$

and for any S with $\{r, s\} \subseteq S \subseteq X$,

$$\pi_r^{\{r,s\}}(S) = \sum_{D\colon \{r,s\} \subseteq D \subseteq S} (-1)^{|S \setminus D|} \frac{p_r(r, X)}{p_r(r, D)}$$

The first equation helps us identify $\pi_r(S)/\pi_r(\{r\})$ for each $S \subseteq X \setminus s$, while the second equation helps us identify $(\pi_r(S \setminus s) + \pi_r(S))/(\pi_r(\{r\}) + \pi_r(\{r, s\}))$ for each S such that $\{r, s\} \subseteq S$. After some normalization we can back out $\pi_r(S)$ for all S. The necessary condition for this procedure to be well defined are that both $\pi_r^r(S) > 0$ and $\pi_r^{\{r,s\}}(S) > 0$. Furthermore, it must be the case that $\pi_r^{\{r,s\}}(S) - \pi_r^r(S) > 0$.

The construction works as follows. For each $S \subseteq X \setminus s$, let

$$\lambda_r^r(S) = \sum_{D\colon r \in D \subseteq S} (-1)^{|S \setminus D|} \frac{1}{p_r(r, D)}$$

and for each S such that $\{r, s\} \subseteq S \subseteq X$, let

$$\lambda_r^{\{r,s\}}(S) = \lambda_r^{\{r,s\}}(\{r, s\}) \sum_{D\colon \{r,s\} \subseteq D \subseteq S} (-1)^{|S \setminus D|} \frac{1}{p_r(r, D)}.$$

where $\lambda_r^{\{r,s\}}(\{r, s\}) > 1$ is a constant to be determined later. The relative weight of each S with respect to r is defined by

$$\lambda_r(S) = \begin{cases} \lambda_r^r(S) & \text{if } r \in S \subseteq X \setminus s, \\ \lambda_r^{\{r,s\}}(S) - \lambda_r^r(S \setminus s) & \text{if } \{r,s\} \subseteq S \subseteq X. \end{cases}$$

$\lambda_r^{\{r,s\}}(\{r,s\})$ is chosen large enough to make sure that $\lambda_r^{\{r,s\}}(S) - \lambda_r^r(S \setminus s) > 0$ for all S such that $\{r,s\} \subseteq S \subseteq X$. This is possible as long as the second part of the equation for $\lambda_r^{\{r,s\}}(S)$ is positive. Finally, let

$$\pi_r(S) = \frac{\lambda_r(S)}{\sum_{D:\, r \in D \subseteq X} \lambda_r(D)}.$$

Notice that unlike the case when r is the worst alternative, there is a degree of freedom which is the choice of $\pi_r(\{r,s\})/\pi_r(\{r\})$.

2.4.2 Representation Theorem

To introduce the axiom, let $\mathcal{D}_r$ be the collection of subsets of X in which r is the dominant alternative. Also let P_r denote the set of alternatives which dominate r in binary comparison. For any (S,r), the odds that an alternative other than the reference point is chosen is defined by $\mathcal{O}_S^r = \frac{1-p_r(r,S)}{p_r(r,S)}$. For any $U \not\ni r$, $\Delta_U \mathcal{O}_S^r = \frac{1-p_r(r,S)}{p_r(r,S)} - \frac{1-p_r(r,S\setminus U)}{p_r(r,S\setminus U)}$. This reflects the change in odds against the reference alternative when the choice set is expanded. The next axiom is related to Increasing Feasible Odds axiom in Brady and Rehbeck (2016).

Axiom 2.4.1. *(Decreasing Odds for the Reference Alternative) For any (S,r) and a collection $\mathcal{U} = \{U_1, \cdots, U_n\}$ such that for all $i \in \{1,\ldots,n\}$, $U_i \subseteq P_r$ and $S \cap U_i \neq \emptyset$,*

$$\Delta_{\mathcal{U}} \mathcal{O}_S^r = \Delta_{U_n} \cdots \Delta_{U_2} \mathcal{O}_S^r - \Delta_{U_n} \cdots \Delta_{U_2} \mathcal{O}_{S\setminus U_1}^r > 0.$$

This axiom states that when the choice set is expanded by introducing alternatives which dominate the reference alternative the odds that the reference alternative is chosen decreases at an increasing rate. Notice that Axiom 2.4.1 is silent on the effect of enlarging a menu with alternatives that are dominated by the reference alternative. Hence, logit model relaxes the condition imposed by IDA which states that the probability that the reference alternative is chosen must be the same when an alternative dominated by it is added to the menu. While IDA and Axiom 2.4.1 are logically independent, from the representation theorem for the fixed independent consideration model we know that IDA together with RIDA implies Axiom 2.4.1.

Observe that Example 2.4.2 does not satisfy Axiom 2.4.1. Indeed, if we choose $\mathcal{U} = \{\{x\}, \{y\}\}$, then

$$\Delta_{\mathcal{U}} \mathcal{O}^{r}_{\{x,y,z\}} = -\frac{7}{9}.$$

Axiom 2.4.1 rules out this example by requiring that the consideration probabilities identified from choices are positive. Going back to the case when r is the second worst alternative in the set and s is the worst alternative, Axiom 2.4.1 is equivalent to the following conditions on choice probabilities:

$$\sum_{D\colon r \in D \subseteq S} (-1)^{|S \setminus D|} \frac{1}{p_r(r, D)} > 0 \quad \text{whenever } S \subseteq X \setminus s,$$
$$\sum_{D\colon \{r,s\} \subseteq D \subseteq S} (-1)^{|S \setminus D|} \frac{1}{p_r(r, D)} > 0 \quad \text{whenever } S \not\subseteq X \setminus s.$$

As noted earlier, these conditions are necessary for the consideration probabilities induced from choice data to be well defined.

The next theorem provides the representation result for logit consideration.

Theorem 2.4.2. *Suppose p_r is an RDRAM. Then p_r has a logit attention representation if and only if it satisfies Ratio Independence of Dominant Alternatives and Decreasing Odds for the Reference Alternative.*

Necessity of Axiom 2.4.1 when r is arbitrary is shown in the Appendix. To prove sufficiency, consideration probabilities are constructed from choices as illustrated in the previous section. Axiom 2.4.1 is used to prove that this is a well defined probability distribution. The construction of consideration probabilities guarantees that the representation always holds for the reference alternative. Hence, this guarantees that the representation holds for both alternatives in binary choice sets. By using induction and RIDA, it is shown that the representation holds for all alternatives and all choice sets.

2.5 Fixed Correlated Consideration

In this section, I study the behavioral implications of the fixed correlated consideration model. In FCC model, each alternative has a fixed reference dependent probability of being considered that does not depend on the choice set. This probability is given by $\sum_{x \in D} \pi_r(D)$. The difference with fixed independent consideration model is that the consideration of alternatives might be correlated. Similar to FIC model, FCC satisfies Irrelevance of Dominated Alternatives. The next examples illustrate that RIDA is not necessarily satisfied in FCC model and IDA by itself is not sufficient for an FCC representation.

Example 2.5.1 (RIDA Violation)**.** *Suppose* $x \succ y \succ z$*,* $\pi_z(\{z\}) = \pi_z(\{x, z\}) = \pi_z(\{y, z\}) = 0.2$ *and* $\pi_z(\{x, y, z\}) = 0.4$*. The implied choice probabilities when* z *is the reference alternative are as below.*

$p_z(\cdot, S)$	$\{x, y, z\}$	$\{x, z\}$	$\{y, z\}$
x	0.6	0.6	
y	0.2		0.6
z	0.2	0.4	0.4

RIDA is violated as $\frac{p_z(y,\{x,y,z\})}{p_z(z,\{x,y,z\})} = 1 \neq \frac{3}{2} = \frac{p_z(y,\{y,z\})}{p_z(z,\{y,z\})}$.

Example 2.5.2 (Insufficiency of IDA)**.** *Suppose* p_r *is an RDRAM and when* z*, the worst alternative in* $\{x, y, z\}$*, is the reference alternative the choice data is as below.*

$p_z(\cdot, S)$	$\{x, y, z\}$	$\{x, z\}$	$\{y, z\}$
x	0.6	0.6	
y	0.2		0.9
z	0.2	0.4	0.1

Notice that IDA is satisfied as $p_z(x, \{x, y, z\}) = p_z(x, \{x, z\})$. If p_z has a fixed correlated attention representation, then it must be the case that $p(z, \{x, y, z\}) = \pi_z(\{z\}) = 0.2$ and $p_z(z, \{y, z\}) = \pi_z(\{z\}) + \pi_z(\{x, z\}) = 0.1$ which is not possible as π_z is a probability distribution.

In Example 2.5.2, we have that $p_z(z, \{y, z\}) < p_z(z, \{x, y, z\})$ which is not allowed in fixed correlated consideration model. Recall that IDA only imposes a condition on observed choices when dominated alternatives are removed from the choice set. Hence, IDA does not impose any condition on $p_z(\cdot, \{y, z\})$. One possible assumption on observed choices is $p_z(z, \{y, z\}) > p_z(z, \{x, y, z\})$. This condition is known as *regularity* (Suppes and Luce, 1965). While this assumption is necessary for FCC representation, it is not sufficient. To see this, modify Example 2.5.2 so that $p_z(z, \{y, z\}) = 0.9$ and $p_z(y, \{y, z\}) = 0.1$. Then regularity on the reference alternative is satisfied. However, if π_z is the FCC representation of p_z, it must be that

D	$\{x, y, z\}$	$\{x, z\}$	$\{y, z\}$	$\{z\}$
$\pi_z(D)$	-0.1	0.7	0.2	0.2

Hence, more structure on observed choices is needed for π_r derived from p_r to be a valid probability distribution.

2.5.1 Identification of Consideration Probabilities

Here I discuss how π_r can be recovered from p_r. First, suppose r is the worst alternative in X. Then if p_r has a fixed correlated consideration representation it must be the case that

$$p_r(r, S) = \sum_{D:\, r \in D \subseteq (X \setminus S) \cup r} \pi_r(D).$$

This can alternatively be written as

$$p_r(r, (X \setminus S) \cup r) = \sum_{D:\, r \in D \subseteq S} \pi_r(D).$$

Hence, by Möbius Inversion,

$$\pi_r(S) = \sum_{D:\, r \in D \subseteq S} (-1)^{|S \setminus D|} p_r(r, (X \setminus D) \cup r).$$

Now suppose r is the second worst alternative. Let s be the worst alternative in X. Notice that Irrelevance of Dominated Alternatives implies

$$p_r(x, S) = p_r(x, S \setminus s)$$

for all S and $x \in S \setminus s$. Hence, unlike logit consideration model, there is no extra information when s is added to the choice set. This implies that we can only identify $\pi_r(D) + \pi_r(D \cup s)$ for each $D \subseteq X \setminus s$ and nothing else. Let $\pi_r^{\{r,s\}}$ be the probability distribution on choice sets $S \supseteq \{r, s\}$ defined by

$$\pi_r^{\{r,s\}}(S) = \pi_r(S) + \pi_r(S \setminus s).$$

Now notice that if $S \supseteq \{r, s\}$, then

$$p_r(r, (X \setminus S) \cup \{r, s\}) = \sum_{D:\, \{r,s\} \subseteq D \subseteq S} \pi_r^{\{r,s\}}(D).$$

Hence, by Möbius Inversion,

$$\begin{aligned}
\pi_r^{\{r,s\}}(S) &= \sum_{D:\, \{r,s\} \subseteq D \subseteq S} (-1)^{|S \setminus D|} p_r(r, (X \setminus D) \cup \{r, s\}) \\
&= \sum_{D:\, \{r,s\} \subseteq D \subseteq S} (-1)^{|S \setminus D|} p_r(r, (X \setminus D) \cup r)
\end{aligned}$$

where the second equality follows if we assume Irrelevance of Dominated Alternatives. Thus, a necessary condition for the representation is positivity of $\pi_r^{\{r,s\}}(S) > 0$ for all $S \supseteq \{r, s\}$.

2.5.2 Representation Theorem

Recall that P_r is the set of alternatives which dominate the reference alternative in binary comparison. Also let $D_r = X \setminus P_r$, i.e., D_r is the largest set in which the reference alternative is dominant. For any (S, r) and $U \not\ni r$, let $\Delta_U p_r(r, S) = p_r(r, S \setminus U) - p_r(r, S)$. This reflects in the change in the probability that an alternative other than the reference alternative is chosen when the set is expanded. The next axiom is similar to Aguiar (2017).

Axiom 2.5.1. *(Decreasing Propensity of Choice for the Reference Alternative) For any (S, r) and a collection $\mathcal{U} = \{U_1, \cdots, U_n\}$ such that for all $i \in \{1, \ldots, n\}$, $U_i \subseteq P_r$ and $S \cap U_i \neq \emptyset$,*

$$\Delta_{\mathcal{U}} p_r(r, S) = \Delta_{U_n} \cdots \Delta_{U_2} p_r(r, S \setminus U_1) - \Delta_{U_n} \cdots \Delta_{U_2} p_r(r, S) > 0.$$

The axiom states that when the choice set is expanded by introducing alternatives which dominate the reference point the probability that the reference point is chosen decreases at a *decreasing* rate. It is illustrative to compare Axiom 2.5.1 with Axiom 2.4.1 assumption in logit model. Recall that Axiom 2.4.1 requires the odds that the reference alternative is chosen to decrease at an *increasing* rate. Since fixed independent consideration satisfies both properties, in FIC model while the odds that the reference alternative is chosen decreases at an increasing rate, the probability that it is chosen decreases at a decreasing rate. To see this in an example, suppose p_r has a FIC representation with γ_r. Let $S' \subsetneq S$ and $x \notin S$ be given. Then a simple calculation shows that $\Delta_{\{x\}} \mathcal{O}^r_{S \cup x} > \Delta_{\{x\}} \mathcal{O}^r_{S' \cup x}$ and $\Delta_{\{x\}} p_r(r, S \cup x) < \Delta_{\{x\}} p_r(r, S' \cup x)$.

Notice that similar to Axiom 2.4.1, Axiom 2.5.1 imposes a structure on observed choices when the choice set is expanded by adding alternatives which dominate the reference point. On the other hand, IDA imposes a structure on observed choices when dominated alternatives are added to the menu. Combining IDA with Axiom 2.5.1 we get a property known as regularity on the reference point: for any $S \supseteq T \ni r$, $p_r(r, T) \geq p_r(r, S)$. This property is not shared by the logit model. In fact, logit consideration imposes no structure on observed choices when an alternative dominated by the reference point is added to the menu.

Going back to the example when r is the second worst alternative in the set, Axiom 2.5.1 is equivalent to the condition that

$$\sum_{D:\, \{r,s\} \subseteq D \subseteq S} (-1)^{|S \setminus D|} p_r(r, (X \setminus D) \cup \{r, s\}) > 0.$$

As discussed earlier, this is a necessary condition for consideration probabilities derived from observed choices to be well defined. The necessity of Axiom 2.5.1 when r is arbitrary is shown in the Appendix.

The next theorem provides a characterization result for fixed correlated consideration.

Theorem 2.5.1. *Suppose p_r is an RDRAM. Then p_r has a fixed correlated consideration representation if and only if it satisfies Irrelevance of Dominated Alternatives and Decreasing Propensity of Choice for the Reference Alternative.*

The proof of sufficiency is provided in three steps. First, I construct consideration probabilities from p_r as described in the earlier section. It is shown that Axiom 2.5.1 implies that this is a valid probability distribution. The next step is to show that the representation holds for the reference alternative, and hence the representation holds for both alternatives in binary choice sets. Finally, induction and IDA are used to show that the representation holds for all alternatives and all choice problems.

2.5.3 Relation to Other Models

An interesting feature of fixed correlated consideration model is that it generalizes Masatlioglu and Ok (2014) to random choice setup. In their model, the decision maker is endowed with a psychological constraint function $Q : X \to \mathcal{X}$ such that given a choice problem (S, r) the DM's deterministic choice is

$$c(S, r) = \operatorname{argmax}(\succ, Q(r) \cap S).$$

This is the limit case of fixed correlated consideration model where π_r assigns probability 1 to the set $Q(r)$.

Another interesting class of models related to fixed correlated consideration is random utility models. Here I illustrate by construction that fixed correlated consideration is a special case of reference dependent random utility model (RDRUM). (For an alternative proof of this result, see Aguiar (2017).) Let $\mathcal{R}$ denote the set of all possible strict rankings on X. In RDRUM, for each reference point, there exists a probability distribution π'_r on $\mathcal{R}$ such that

$$p_r(x,S) = \sum_{\succ \in \mathcal{R}:\, x = \mathrm{argmax}(\succ, S)} \pi'_r(\succ).$$

To see how RDRUM generalizes fixed correlated consideration, suppose p_r has an FCC representation with $(\succ, \pi_r)$ where π_r is a probability on $\mathcal{X}$. We can construct an RDRUM representation of p_r with π'_r as follows. First, for any strict preference $\succ'$, let $\succ'_D$ be the restriction of $\succ'$ on D. Let $m_\succ : \mathcal{X} \to \mathcal{R}$ be a function defined by

$$m_\succ(D) = \succ' \text{ such that } \succ'_D = \succ_D,\ \succ'_{X\setminus D} = \succ_{X\setminus D} \text{ and } x \succ' y \text{ if } x \in D, y \in X \setminus D.$$

Now for any $\succ' \in \mathcal{R}$, let $\pi'_r(\succ') = \sum_{D:\, m_\succ(D) = \succ'} \pi_r(D)$. Then π'_r defined as such is an RDRUM representation of p_r. Since fixed independent consideration is a special case of fixed correlated consideration, it is also a special case of RDRUM. However, logit consideration model and random utility models are independent.

2.6 Conclusion

In this paper, I consider reference dependent random choice data. I provide novel characterization results for three commonly used random attention models. It is shown fixed independent consideration model can be characterized by two key properties: irrelevance of dominated alternatives and ratio independence of dominant alternatives. Logit consideration relaxes the first property while fixed correlated consideration relaxes the second property. Hence, the intersection of logit and fixed correlated consideration models is exactly fixed independent consideration. I illustrate how attention parameters can be partially identified in all these models.

2.7 Appendix

Proof of Theorem 2.2.1

For any $x \neq y$, let $x \succ y$ if $p_x(x, \{x, y\}) = 1$. Dominant Alternative implies that $\succ$ is total and asymmetric. To see that $\succ$ is transitive, suppose $x \succ y \succ z$. Then $p_x(x, \{x, y\}) = 1$ and $p_y(y, \{y, z\}) = 1$. Consistency of Dominance implies that $p_y(x, \{x, y, z\}) > 0$ and $p_z(y, \{x, y, z\}) > 0$. Since $\{x, y, z\}$ has a dominant alternative, it must be x. Hence, $p_x(x, \{x, y, z\}) = 1$. But then we cannot have $p_z(z, \{x, z\}) = 1$ as this would imply $p_x(z, \{x, y, z\}) > 0$, a contradiction. Hence, $p_x(x, \{x, z\}) = 1$ and $x \succ z$ follows.

Now for any x, let D_x be the largest set in which x is the dominant alternative. Now given a choice problem (S, r) and $x \in S$ with $p_r(x, S) > 0$, let $\mu_r(D, S) > 0$ be such that

$$p_r(x, S) = \sum_{D\colon \{r,x\} \subseteq D \subseteq D_x \cap S} \mu_r(D, S).$$

It is easy to see that this results in the desired representation.

Proof of Theorem 2.3.1

Let $x \succ y$ whenever $p_y(x, \{x, y\}) > 0$. Since p_r is an RDRAM, $\succ$ is a linear order on X. The next lemma provides a relationship between $p_r(y, S)$ and $p_r(y, S \setminus x)$ where x is the dominant alternative in S.

Lemma 2.7.1. *Suppose p_r satisfies IDA and RIDA. If x is the dominant alternative in S, then for any $y \in S \setminus x$,*

$$p_r(y, S) = p_r(y, S \setminus x) p_r(r, \{x, r\}).$$

Proof. RIDA implies that there exists a constant $\kappa(S)$ such that $p_r(y, S) = \kappa(S)p_r(y, S \setminus x)$ for all $y \in S \setminus x$. By adding over all $y \in S \setminus x$, we get that $\sum_{y \in S \setminus x} p_r(y, S) = \kappa(S)$, or alternatively, $1 - p_r(x, S) = \kappa(S)$. Now IDA implies that $1 - p_r(x, S) = 1 - p_r(x, \{r, x\}) = p_r(r, \{x, r\})$. Hence, $p_r(y, S) = p_r(y, S \setminus x)p_r(r, \{x, r\})$. □

The next step is to define $\gamma_r(x)$. Let $\gamma_r(x) = \max\{p_r(x, \{r, x\}), p_x(r, \{r, x\})\}$. Since p_r is an RDRAM, exactly one of the terms is strictly between zero and one so that $\gamma_r(x) \in (0, 1)$. The consideration set probabilities are defined by

$$\mu_r(D, S) = \prod_{x \in D} \gamma_r(x) \prod_{y \in S \setminus D} (1 - \gamma_r(y)).$$

Lemma 2.7.2. *For any* (S, r) *and* $D \subseteq S \setminus x$*,*

$$\mu_r(D, S) = \mu_r(D, S \setminus x)\mu_r(\{r\}, \{r, x\}).$$

Proof. This claim follows from simple algebra and definitions.

$$\begin{aligned}\mu_r(D, S) &= \prod_{y \in D} \gamma_r(y) \prod_{z \in S \setminus D} (1 - \gamma_r(z)) = \prod_{y \in D} \gamma_r(y) \prod_{z \in S \setminus (D \cup x)} (1 - \gamma_r(z))(1 - \gamma_r(x)) \\ &= \mu_r(D, S \setminus x)\mu_r(\{r\}, \{r, x\})\end{aligned}$$

□

Lemma 2.7.3. *For any* (S, r) *and* $x \in S$*,*

$$p_r(x, S) = \sum_{D \colon x \text{ is } \succ \text{ best in } D} \mu_r(D, S)$$

Proof. The proof is by induction. First, suppose $S = \{x, y\}$. Without loss of generality, assume $x \succ y$. Then $p_x(x, \{x, y\}) = 1 = \mu_x(\{x\}, \{x, y\}) + \mu_x(\{x, y\}, \{x, y\})$. Also, by definition, $p_y(x, \{x, y\}) = \gamma_y(x) = \mu_y(\{x, y\}, \{x, y\})$ since $\gamma_y(y) = 1$.

Now suppose the claim holds for all choice sets with cardinality less than k. Let $|S| = k$ and $r \in S$ be given. Suppose y is the dominant alternative in S. If $y = r$, then the claim is trivial since this implies $p_r(r, S) = 1$ and $p_r(x, S) = 0$ for all $x \in S \setminus r$. Hence, suppose $y \neq r$ and let $x \neq y$ be given. Then,

$$\begin{aligned} p_r(x,S) &= p_r(x, S\setminus y) p_r(r, \{r,y\}) \\ &= \Big(\sum_{D\colon x \text{ is } \succ \text{ best in } D \subseteq S\setminus y} \mu_r(D, S\setminus y) \Big) \mu_r(r, \{r,y\}) \\ &= \sum_{D\colon x \text{ is } \succ \text{ best in } D \subseteq S\setminus y} \mu_r(D, S) \\ &= \sum_{D\colon x \text{ is } \succ \text{ best in } D} \mu_r(D, S) \end{aligned}$$

where the first equality follows from the first lemma, the second equality follows from induction, the third equality follows from the previous lemma, and the last equality follows from the fact that $y \succ x$. Finally, since $p_r(y, S) = 1 - \sum_{x \in S\setminus y} p_r(x, S)$, the representation holds for all $x \in S$. Since the reference point was arbitrary, the claim holds for all choice problems (S, r). $\square$

Proof of Theorem 2.4.2

Suppose p_r has a logit attention representation. The necessity of RIDA is illustrated in the main text. Here I show the necessity of Axiom 2.4.1.

Lemma 2.7.4. *Suppose p_r has a logit attention representation $(\succ, \pi_r)$. Then p_r satisfies Axiom 2.4.1.*

Proof. Recall that P_r denotes all alternatives which dominate r in binary comparison (i.e., $s \in P_r \Leftrightarrow p_s(s, \{r, s\}) = 1$). For any S,

$$\mathcal{O}^r_S = \frac{1 - p_r(r,S)}{p_r(r,S)} = \frac{\sum_{D\colon r\in D\subseteq S} \pi_r(D) - \sum_{D\colon r\in D\subseteq S\setminus P_r} \pi_r(D)}{\sum_{D\colon r\in D\subseteq S\setminus P_r} \pi_r(D)} = \frac{\sum_{D\colon r\in D\subseteq S} \pi_r(D)}{\sum_{D\colon r\in D\subseteq S\setminus P_r} \pi_r(D)} - 1.$$

so that for any $U \subseteq P_r$,

$$\Delta_U \mathcal{O}^r_S = \frac{1-p_r(r,S)}{p_r(r,S)} - \frac{1-p_r(r,S\setminus U)}{p_r(r,S\setminus U)} = \frac{\sum_{D:\, r\in D\subseteq S}\pi_r(D) - \sum_{D:\, r\in D\subseteq S\setminus U}\pi_r(D)}{\sum_{D:\, r\in D\subseteq S\setminus P_r}\pi_r(D)}$$
$$= \frac{\sum_{D:\, r\in D\subseteq S \text{ and } D\cap U\neq\emptyset}\pi_r(D)}{\sum_{D:\, r\in D\subseteq S\setminus P_r}\pi_r(D)}.$$

Similarly, for $\mathcal{U} = \{U_1, \dots, U_n\}$ such that for all $i \in \{1, \dots, n\}$, $U_i \subseteq P_r$ and $S \cap U_i \neq \emptyset$,

$$\Delta_{\mathcal{U}} \mathcal{O}^r_S = \Delta_{U_n}\cdots\Delta_{U_2}\mathcal{O}^r_S - \Delta_{U_n}\cdots\Delta_{U_2}\mathcal{O}^r_{S\setminus U_1} = \frac{\sum_{D:\, r\in D\subseteq S \text{ and } \forall i,\, D\cap U_i\neq\emptyset}\pi_r(D)}{\sum_{D:\, r\in D\subseteq S\setminus P_r}\pi_r(D)}$$
$$\geq \frac{\pi_r(S)}{\sum_{D:\, r\in D\subseteq S\setminus P_r}\pi_r(D)} > 0.$$

This concludes the proof that Axiom 2.4.1 is necessary for logit attention representation. □

The next step is to show that RIDA and Axiom 2.4.1 are sufficient for logit consideration. First, define $x \succ y$ if $p_x(x, \{x, y\}) = 1$. As before, $\succ$ is a linear order on X. To construct π_r from p_r, let $\lambda_r(\{r\}) = 1$. Let $\mathcal{D}_r$ be the collection of subsets of X which have r as the dominant alternative. For $T \in \mathcal{D}_r$ such that $T \neq \{r\}$, $\lambda_r(T)$ are positive constants to be determined later. Now for each $T \in \mathcal{D}_r$, let

$$\lambda^T_r(T) = \sum_{D:\, r\in D\subseteq T}\lambda_r(D)$$

and for each S such that $T \subseteq S \subseteq T \cup P_r$

$$\lambda^T_r(S) = \lambda^T_r(T)\sum_{D:\, T\subseteq D\subseteq S}(-1)^{|S\setminus D|}\frac{1}{p_r(r,D)}.$$

The next lemma shows that if p_r satisfies Axiom 2.4.1, then $\lambda^T_r(S) > 0$.

Lemma 2.7.5. *Suppose p_r is an RDRAM and it satisfies Axiom 2.4.1. Then for any $T \in \mathcal{D}_r$, and S such that $T \subseteq S \subseteq T \cup P_r$, $\lambda^T_r(S) > 0$.*

Proof. Given the definition of $\lambda_r^T(S)$ and using Möbius Inversion, we get that

$$p_r(r,S) = \frac{\lambda_r^T(T)}{\sum_{D\colon T\subseteq D\subseteq S}\lambda_r^T(S)}.$$

Now let $\mathcal{U} = \{\{x_i\} | \, x_i \in S \setminus T\}$. Then a simple calculation shows that

$$0 < \Delta_{\mathcal{U}}\mathcal{O}_S^r = \Delta_{U_n}\cdots\Delta_{U_2}\mathcal{O}_S^r - \Delta_{U_n}\cdots\Delta_{U_2}\mathcal{O}_{S\setminus U_1}^r = \frac{\sum_{D\colon T\subseteq D\subseteq S \text{ and } \forall i,\, D\cap U_i\neq\emptyset}\lambda_r^T(D)}{\lambda_r^T(T)} = \frac{\lambda_r^T(S)}{\lambda_r^T(T)}.$$

Since the denominator is positive by definition, the numerator is also positive. ☐

Now let $\lambda_r(S) = \lambda_r^r(S)$ for S such that $S \setminus P_r = \{r\}$. Suppose $\lambda_r(S)$ is defined for all S with $|S \setminus P_r| < k$ and let S be such that $|S \setminus P_r| = k$. Then $\lambda_r(S)$ is given by

$$\lambda_r(S) = \lambda_r^{S\setminus P_r}(S) - \sum_{D\colon r\in D\subsetneq S\setminus P_r}\lambda_r(D\cup(S\cap P_r)).$$

Since by the previous lemma the second part in the equation for $\lambda_r^{S\setminus P_r}(S)$ is positive, by choosing $\lambda_r(S \setminus P_r)$ large enough we can make sure that $\lambda_r(S) > 0$.

Lemma 2.7.6. *For any $T \in \mathcal{D}_r$ and S such that $T \subseteq S \subseteq T \cup P_r$,*

$$\lambda_r^T(S) = \sum_{(S\setminus T)\cup r\subseteq D\subseteq S}\lambda_r(D).$$

Proof. The claim follows from the definition if $S = T$. For $T \subsetneq S$, by definition,

$$\begin{aligned}\lambda_r^T(S) &= \lambda_r(S) + \sum_{r\in D\subsetneq T}\lambda_r(D\cup(S\setminus T))\\ &= \sum_{D\colon (S\setminus T)\cup r\subseteq D\subseteq S}\lambda_r(D).\end{aligned}$$

☐

Lemma 2.7.7. *Suppose p_r is an RDRAM and it satisfies RIDA and Axiom 2.4.1. Then for any*

$r \in X$ and $x \in S$,

$$p_r(x,S) = \sum_{D\colon x \text{ is } \succ \text{ best in } D} \frac{\lambda_r(D)}{\sum_{D'\colon r\in D'\subseteq S} \lambda_r(D')}.$$

Proof. The proof is by induction. Let $r \in X$ be given. First, I show that the representation holds for binary choice sets. If $p_r(x,\{r,x\}) = 0$, then $r \succ x$ and the representation holds trivially. Suppose $p_r(x,\{r,x\}) > 0$. By construction, $\lambda_r(\{r\}) = 1$ and $\lambda_r(\{r,x\}) = \frac{1}{p_r(r,\{r,x\})} - 1$. Hence,

$$\frac{\lambda_r(\{r,x\})}{\lambda_r(\{r\}) + \lambda_r(\{r,x\})} = 1 - p_r(r,\{r,x\}) = p_r(x,\{r,x\}).$$

The next step is to show that the claim holds for all S whenever $x = r$. Let $T = S \setminus P_r$. Notice that Möbius Inversion implies

$$\frac{\lambda_r^T(T)}{p_r(r,S)} = \sum_{D\colon T\subseteq D\subseteq S} \lambda_r^T(D),$$

and hence

$$p_r(r,S) = \frac{\lambda_r^T(T)}{\sum_{D\colon T\subseteq D\subseteq S} \lambda_r^T(D)} = \frac{\sum_{D'\colon r\in D'\subseteq T} \lambda_r(D')}{\sum_{D\colon T\subseteq D\subseteq S}\sum_{D''\colon (D\setminus T)\cup r\subseteq D''\subseteq D} \lambda_r(D'')} = \frac{\sum_{D'\colon r\in D'\subseteq T} \lambda_r(D')}{\sum_{D''\colon r\in D''\subseteq S} \lambda_r(D'')}$$

where the second equality follows from the previous lemma. Since r is the $\succ$ best alternative only in subsets of T, this proves the lemma whenever $x = r$.

Now suppose the lemma holds for all S with cardinality less than k and let S with $|S| = k$ be given. If x is not chosen with positive probability, the representation is trivial. So suppose x is chosen with positive probability. If x is the only alternative other than r chosen with positive probability, then the representation is trivial again as $p_r(x,S) = 1 - p_r(r,S)$ and the representation holds for the reference alternative. So let $z \neq x$ be the dominant alternative in S. RIDA implies that

$$\frac{p_r(x,S)}{p_r(r,S)} = \frac{p_r(x,S\setminus z)}{p_r(r,S\setminus z)}.$$

Hence, by using induction argument, the fact that z is the dominant alternative in S, and that the representation holds whenever $x = r$, we get

$$p_r(x,S) = \frac{p_r(x, S\setminus z)}{p_r(r, S\setminus z)} p_r(r,S) = \frac{\frac{\sum_{D:\, x \text{ is } \succ \text{ best in } D\subseteq S\setminus z} \lambda_r(D)}{\sum_{r\in D'\subseteq S\setminus z}\lambda_r(D')}}{\frac{\sum_{D:\, r \text{ is } \succ \text{ best in } D\subseteq S\setminus z} \lambda_r(D)}{\sum_{r\in D'\subseteq S\setminus z}\lambda_r(D')}} \frac{\sum_{D:\, r \text{ is } \succ \text{ best in } D\subseteq S} \lambda_r(D)}{\sum_{r\in D'\subseteq S}\lambda_r(D')}$$
$$= \frac{\sum_{D:\, x \text{ is } \succ \text{ best in } D\subseteq S} \lambda_r(D)}{\sum_{D:\, r \text{ is } \succ \text{ best in } D\subseteq S} \lambda_r(D)} \frac{\sum_{D:\, r \text{ is } \succ \text{ best in } D\subseteq S} \lambda_r(D)}{\sum_{r\in D'\subseteq S}\lambda_r(D')}$$
$$= \sum_{D:\, x \text{ is } \succ \text{ best in } D\subseteq S} \frac{\lambda_r(D)}{\sum_{r\in D'\subseteq S}\lambda_r(D')}$$

as desired. Finally, the claim holds for z as $p_r(z,S) = 1 - \sum_{y\in S\setminus z} p_r(y,S)$. □

Define a probability distribution π_r on sets which contain r by

$$\pi_r(S) = \frac{\lambda_r(S)}{\sum_{r\in D\subseteq X}\lambda_r(D)}.$$

The previous lemma shows that $(\succ, \pi_r)$ represents p_r. Since (S,r) was arbitrary, this concludes the proof of the theorem.

Proof of Theorem 2.5.1

Suppose p_r has a fixed correlated consideration. The necessity of IDA is shown in the main text. Here I prove the necessity of Axiom 2.5.1.

Lemma 2.7.8. *Suppose p_r has an FCC representation with π_r. Then p_r satisfies Axiom 2.5.1.*

Proof. Let D_r to be the largest set in which r is the dominant alternative, i.e., $D_r = X\setminus P_r$. First, notice that for any (S,r) and any $U\subseteq P_r$ such that $S\cap U \neq \emptyset$,

$$\Delta_U p_r(r,S) = p_r(r, S\setminus U) - p_r(r,S) = \sum_{D:\, r\in D\subseteq [X\setminus(S\setminus U)]\cup D_r} \pi_r(D) - \sum_{D:\, r\in D\subseteq [X\setminus S]\cup D_r} \pi_r(D)$$
$$= \sum_{D:\, r\in D\subseteq [X\setminus(S\setminus U)]\cup D_r \text{ and } D\cap(S\cap U)\neq\emptyset} \pi_r(D) \geq \pi_r((S\cap U)\cup r) > 0.$$

Similarly, for any collection $\mathcal{U} = \{U_1, \ldots, U_n\}$ such that for all $i \in \{1, \ldots, n\}$, $U_i \subseteq P_r$ and $S \cap U_i \neq \emptyset$,

$$\Delta_{\mathcal{U}} p_r(r, S) = \sum_{D:\, r \in D \subseteq [X \setminus (S \setminus (\cup_{U_i \in \mathcal{U}} U_i))] \cup D_r \text{ and } \forall i,\, D \cap (S \cap U_i) \neq \emptyset} \pi_r(D) \geq \pi_r((S \cap (\cup_{U_i \in \mathcal{U}} U_i)) \cup D_r) > 0.$$

$\square$

I will now show that if p_r is an RDRAM, then IDA and Axiom 2.5.1 are sufficient for an FCC representation. As before, we let $x \succ y$ iff $p_x(x, \{x, y\}) = 1$. Let $r \in X$ be given. Now for any $S \supseteq D_r$, let

$$\lambda_r(S) = \sum_{D:\, D_r \subseteq D \subseteq S} (-1)^{|S \setminus D|} p_r(r, (X \setminus D) \cup D_r).$$

The next lemma shows that if p_r satisfies Axiom 2.5.1, then $\lambda_r(S) > 0$ for any choice problem (S, r) such that $S \supseteq D_r$.

Lemma 2.7.9. *Suppose p_r is an RDRAM and it satisfies Axiom 2.5.1. Then $\lambda_r(S)$ defined as above is strictly positive for all (S, r) with $S \supseteq D_r$.*

Proof. Given the definition of λ_r, Möbius Inversion implies that

$$p_r(r, S) = \sum_{D:\, D_r \subseteq D \subseteq (X \setminus S) \cup D_r} \lambda_r(D).$$

Let $\mathcal{U} = \{\{x_i\} | \, x_i \in S \cap P_r\}$. By assumption, $\Delta_{\mathcal{U}} p_r(r, X) > 0$. Now notice that

$$0 < \Delta_{\mathcal{U}} p_r(r, X) = \sum_{D:\, D_r \subseteq D \subseteq [X \setminus (X \setminus (\cup_{U_i \in \mathcal{U}} U_i))] \cup D_r \text{ and } \forall i,\, D \cap (X \cap U_i) \neq \emptyset} \lambda_r(D) = \lambda_r(S)$$

as desired. $\square$

Notice that Möbius inversion implies

$$p_r(r, D_r) = \sum_{D:\, D_r \subseteq D \subseteq X} \lambda_r(D).$$

Since $p_r(r, D_r) = 1$, this together with the previous lemma imply that λ_r is a probability distribution on sets $S \supseteq D_r$. Now choose π_r over $\mathcal{X}$ containing r which satisfies $\pi_r(D) > 0$ for each $D \in \mathcal{X}$ and for any $S \supseteq D_r$

$$\lambda_r(S) = \sum_{D\colon (S \setminus D_r) \cup r \subseteq D \subseteq S} \pi_r(D).$$

Lemma 2.7.10. *Suppose p_r is an RDRAM and it satisfies IDA and Axiom 2.5.1. Then for any (S, r) and $x \in S$,*

$$p_r(x, S) = \sum_{D\colon x \text{ is } \succ \text{ best in } D \cap S} \pi_r(D).$$

Proof. First, notice that for any $S \supseteq D_r$,

$$p_r(r, S) = \sum_{D\colon D_r \subseteq D \subseteq (X \setminus S) \cup D_r} \lambda_r(D) = \sum_{D\colon r \in D \subseteq (X \setminus S) \cup D_r} \pi_r(D) = \sum_{D\colon r \in D \cap S \subseteq D_r} \pi_r(D).$$

On the other hand, if $S \not\supseteq D_r$, then Irrelevance of Dominated Alternatives implies that

$$p_r(r, S) = p_r(r, S \cup D_r) = \sum_{D\colon r \in D \cap (S \cup D_r) \subseteq D_r} \pi_r(D) = \sum_{D\colon r \in D \cap S \subseteq D_r} \pi_r(D).$$

Hence, the representation holds for the reference alternative. Now let $x \succ r$ be given. Since the representation holds for $p_r(r, \{x, r\})$ it also holds for $p_r(x, \{x, r\})$. Suppose the representation holds for all sets with cardinality less than k. Let S with $|S| = k$ be given. If S contains one alternative which dominates r, the representation holds by induction and irrelevance of dominated alternatives. Suppose there are at least two alternatives which dominate r. Let $z \in S$ be an alternative which dominates r but is dominated by all other alternatives for which $p_r(x, S) > 0$. Let $x \notin \{r, z\}$ with $p_r(x, S) > 0$ be given. By Irrelevance of Dominated Alternatives,

$$p_r(x, S) = p_r(x, S \setminus z).$$

By induction argument,

$$p_r(x, S \setminus z) = \sum_{D:\, x \text{ is } \succ \text{ best in } D \cap S \setminus z} \pi_r(D).$$

Finally, since $x \succ z$, x is $\succ$ best in $D \cap S$ if and only if it is $\succ$ best in $D \cap S \setminus z$, and hence

$$p_r(x, S) = p_r(x, S \setminus z) = \sum_{D:\, x \text{ is } \succ \text{ best in } D \cap S} \pi_r(D).$$

To conclude the proof, notice that $p_r(z, S) = 1 - \sum_{x \in S \setminus z} p_r(x, S)$. $\square$

CHAPTER 3

Decision Making within a Product Network[1]

3.1 Introduction

Consider a Netflix consumer who is searching for a movie. First, she looks up a particular movie recommended by a friend. Then, Netflix recommends several other films. These recommendations help customers who face the overwhelming task of finding the best movie among a huge catalog. The recommendations form a *product network* in which a large number of movies are linked to one another. Each movie is a node in the product network, and a recommendation represents a link between two movies. While one of the first e-commerce websites to introduce product networks was Netflix, nowadays almost every e-commerce site offers product networks.[2]

In marketing, researchers recognize the importance of product networks in decision-making (Hoffman and Novak, 1996; Mandel and Johnson, 2002; Bucklin and Sismeiro, 2003). Moe (2003) claims that "[e]normous potential exists in studying an individual's behavior as they navigate from page to page." This navigation is the key difference between the classical decision-making examined in economics and decision-making within product networks. Shopping within a product network is analogous to walking down the aisles of a supermarket. As in traditional supermarkets, each product has a virtual "shelf position" in the product network, which immensely affects demand for that product (Johnson et al., 2004; Oestreicher-Singer and Sundararajan, 2012). By improving their knowledge of consumer decision-making, firms can improve their marketing strate-

[1] This chapter is coauthored with Yusufcan Masatlioglu.

[2] As of 2016, Netflix is the world's leading Internet television network with over 93 million members in over 190 countries enjoying over 125 million hours of TV shows and movies per day.

gies. Hence, understanding how a product network shapes consumers' search has become crucial for businesses. Surprisingly, theoretical work on decision-making within a product network is very limited.

In this paper, we investigate decision-making in this new environment where consumers encounter products in a product network. We consider a customer who picks her most preferred option from the alternatives she can discover (not from the entire feasible set). By studying the observed choice behavior of consumers, our aim is to infer not only their preferences but also the product network they face. We propose an identification strategy to find out when and how one can deduce both consumer's preferences and the perceived product network she faces from observed choices.[3] Our identification will help the analyst to pinpoint: (i) which products belong to the same subcategory (i.e., connected components)? (ii) which products are more relevant in linking popular products to niche products? (iii) which products are more likely to trigger sales of a particular product? The ability to answer these questions could be of practical importance for firms, regulators, and policy makers.

Throughout this paper, we assume that the product network of a consumer is not observable, as well as the preferences of the consumer. There are three important reasons behind this assumption. First, recommendation systems offer personalized suggestions to match customers with the most appropriate products. Hence, two Netflix customers might face two different product networks because of their unobservable characteristics. Second, there is a distinction between perceived versus exogenous networks. Since the actual search is influenced by different factors such as brand familiarity (Baker et al., 1986), packaging (Garber, 1995), and color (Aaker, 1997), the perceived network may be different from the product network exogenously provided as in Netflix. The third is that the perceived product network could solely be an outcome of the consumer's memory and the associations she makes between different alternatives. Indeed, cognitive psychologists have illustrated that our memory is organized in an associative network where nodes represent

[3]The variation in the set of available options is the key for our analysis. Such variations are commonly observed in the real world. For example, Netflix's website removes "Watch Instantly" titles that are not currently available due to different factors, such as delayed license renewals, quality issues, and server technical difficulties. Hence, customers will face different choice sets depending on availability.

products, and links represent connections between products (Collins and Loftus, 1975; Meyer and Schvaneveldt, 1976; Anderson, 1983; Hunt and Ellis, 1999; Anderson and Bower, 2014; Gentner and Stevens, 2014). The fact that the links exist in associative memory does not necessarily mean that the links are activated in reality. Several factors affect the activation of a link such as shape, color, and smell (McCracken and Macklin, 1998). Hence, the presence of a particular product might trigger associated memories and help the consumer recall other associated products.

One of the fundamental assumptions in our model is that the decision maker has well-defined preferences unaffected by the search she undertakes. This assumption has been used in empirical research in marketing (Kim et al., 2010; Koulayev, 2010; Chen and Yao, 2016; Honka and Chintagunta, 2016). Bronnenberg et al. (2016) provide supporting evidence for the assumption by showing that the valuation of an alternative is the same during the search as well as at the time of purchase. Similarly, Reutskaja et al. (2011) provide experimental evidence for the claim that subjects are good at optimizing within the set of options they have explored.

Every search in a network starts from one of the available alternatives. We call this alternative the starting point in the consumer's search.[4] In every stage of the search, the consumer will "click on" all the recommendations that appear in her perceived network. Hence, in the baseline model, there are no cognitive limitations other than the one induced by her perceived product network. Our consumer will consider all the goods that are reachable from the starting point. However, as opposed to the standard theory, she might not discover all available alternatives if the product network is sparse.[5]

Our network model is capable of accommodating two seemingly anomalous types of choice behavior. First, given a fixed menu, different starting points can result in different choices (starting point dependence). Moreover, given a fixed starting point, removing irrelevant alternatives from a menu can affect choices (a violation of the weak axiom of revealed preference). While our

[4]Different starting points may result in different consideration sets, and hence different choices (Bronnenberg et al., 2016). Our initial analysis assumes that we, as the analyst, observe both choices and corresponding starting points. Nevertheless, it is conceivable that starting points will not be observable in some situations. We study the case of unobserved starting points in the Appendix.

[5]The standard model is a special case of our baseline model when the product network is complete (i.e., all the alternatives in the product network are linked).

model helps explain these types of puzzling choice behavior, it also offers distinctive predictions so that we can gain new insights. One of the predictions of the model is that the introduction of a new alternative never diminishes the customer's welfare. Hence, the model generates falsifiable predictions even when the product network are not observable.

Section 3.2 illustrates how we can infer the preference and the product network of a customer only from observed choices. Identifying preferences is particularly important for welfare analysis. As Bernheim and Rangel (2009) point out, it is typically difficult to identify preference from boundedly rational behavior. Our first result shows that the product network is uniquely identified. However, identifying preferences is tricky. We first provide a list of choice patterns which reveal the consumer's preference. Hence, even with limited data, it is possible to identify the preferences. This is crucial for empirical studies when the data is limited. We then show that if we have access to complete choice data, all the patterns that reveal preferences can simply be summarized by the following condition: x is revealed to be preferred to y if and only if x is chosen when y is the starting point.

Our identification strategy relies on the underlying choice procedure where the decision maker maximizes her preference within the alternatives she can reach. It is natural to question the falsifiability of our model. To answer this question, in Section 3.2, we identify the class of choice behaviors that are consistent with our baseline model. We show that choice data is compatible with the network model if and only if data satisfies four simple behavioral postulates. The first one, Starting Point Contraction, says that if the starting point is chosen in a choice set, then it must be chosen from any subset of it as long as it is still the starting point. The second one, Replacement, states that if an alternative other than the starting point is chosen in a choice set, then the original starting point and the chosen alternative will induce the same choices in larger sets. To introduce the next axiom, we call an alternative *influential* in a choice problem, if the removal of this alternative causes a choice reversal. The third axiom, Asymmetry of Influence, states that if y is influential in a choice problem when x is the starting point, then x cannot be influential in the same choice problem with y as the starting point. Lastly, two alternatives are called *distant* in a choice set if for any proper subset of this choice set replacing one alternative with the other alternative as the starting point affects the choice. The last axiom, Irrelevant if Distant, states that if two alternatives are distant in a choice set, then either of these alternatives cannot be influential in this choice problem when the other alternative is the starting point. The key feature of our approach is that our assumptions are stated in terms of choice experiments, and therefore a revealed preference type analysis can be used to test our model.

In Section 3.3, we introduce bounded rationality in our framework. In our baseline model, given the perceived network, the consumer explores all available options which are connected to the starting point. However, many studies show that consumers indeed engage in a very limited search (Johnson et al., 2004; Keane et al., 2008; Goeree, 2008; Kim et al., 2010). We consider a model in which the consumer terminates her search after a number of steps due to limitations such as time pressure or limited cognitive capacity. This model reduces to our baseline model if the number of steps, denoted by K, is larger than the number of products. In the extreme case ($K = 2$), the consumer only considers the starting point and the alternatives which are linked to the starting point. While this model is behaviorally distinct from our baseline model, it coincides with the status quo bias model of Masatlioglu and Ok (2014).

To study K-step network choice, we first provide a characterization for any fixed K. Even in the presence of bounded rationality, three of our original four axioms (Starting Point Contraction, Asymmetry of Influence, and Irrelevant if Distant) hold in K-step network choice model for arbitrary K. On the other hand, Replacement holds if the choice set does not have more than K alternatives. In addition, we introduce a new axiom, K-Reduction, which says that if an alternative is chosen in a choice problem, then there must exist a choice problem with at most K alternatives such that the same alternative is chosen. This is a necessary condition in this model since the final choice must be reachable in at most K steps. The baseline model violates this axiom if K is small enough.

Our limited search model features an interesting relationship between K-Reduction and K-Replacement. While K-Reduction relates *different* choice problems with the *same* starting point,

K-Replacement connects the *same* choice problem with *different* starting points. First, as K decreases, the power of K-Reduction is amplified. While the axiom has no bite when K is big enough, it is very powerful when $K = 2$: removing irrelevant alternatives can never affect choice behavior. Hence, if a consumer follows K-step network choice, given a fixed starting point, fewer violations of "rationality" will be observed as K gets smaller. On the other hand, as K decreases, the power of K-Replacement diminishes. If K is big enough, then replacing the starting point x with another alternative y does not affect the choice as long as x is abandoned for y in a binary comparison. If $K = 2$, then changing the starting point might affect the final choice drastically. Hence, a consumer who follows K-step network choice will be more influenced by the starting point as K gets smaller. To summarize, K-step network choice model relates more search (or a higher K) with fewer observations of starting point dependence and more violations of the weak axiom of revealed preference for a fixed starting point.

Theoretical work on decision making within product networks is limited. The closest paper on the topic that we know of is Masatlioglu and Nakajima (2013). They provide a framework to study behavioral search by utilizing the idea of consideration sets.[6] Their baseline model is quite general and makes very limited predictions. They also consider a particular case which can be represented as a network. However, each model follows completely different choice procedures and they are behaviorally distinct. Their model satisfies *irrelevance of inferior alternatives* and *expansion*, both of which we relax. Irrelevance of inferior alternatives states that adding an alternative that is inferior to the starting point cannot affect the choice. Expansion states that if a starting point x dominates both y and z in binary comparisons, then the starting point x will still be chosen when y and z are presented simultaneously. Hence, on this dimension, their model is more restrictive than ours. On the other hand, their model violates our Replacement axiom.

This paper also contributes to a few branches of decision theory, such as reference dependent choice, limited attention, and search. The other most closely related papers are Tversky and

[6] Caplin and Dean (2011) also study search by employing the revealed preference approach. They assume that an outside observer can view the entire path followed during the search. The main difference is that, in our model, the path is not an input rather an output.

Kahneman (1991); Masatlioglu and Ok (2005, 2014), and Dean et al. (2017) (reference dependent choice); Manzini and Mariotti (2007, 2012); Masatlioglu et al. (2012), and Cherepanov et al. (2013) (limited attention); and Caplin and Dean (2011) (search).

3.2 Model

The marketplace consists of a finite set of alternatives, denoted by X. Our consumer's choice problem can be summarized as a pair (S, x) where S is the set of available products at the time of decision and the consumer starts searching from an alternative x in S. Examples of a starting point include (i) the last purchase or status quo, (ii) a product advertised to the consumer, or (iii) a recommendation from someone in the consumer's social network (Masatlioglu and Nakajima, 2013). With the explosion of data mining technologies, observability of such data is now plausible.[7] The choice behavior of the consumer is summarized by a choice function c which assigns a single element to each extended choice problem (S, x) where $c(S, x) \in S$.

A product network $\mathcal{N}$ is a binary relation on X where $(x, y) \in \mathcal{N}$ represents a direct link between x and y. Our interpretation of $(x, y) \in \mathcal{N}$ is that y belongs to the set of alternatives recommended to a consumer who considers x. We assume that $\mathcal{N}$ is reflexive and symmetric.[8] Intuitively, this implies that a consumer must always consider the starting point of search. In addition, if a consumer with the starting point x considers y, then she must also consider x when her starting point is y.[9]

We assume that the analyst has no prior knowledge about the product network and tries to pin it down from observed choice behavior. There are two reasons for this assumption. First, it is possible that different consumers face different product networks. This can be due to either

[7] For the discussion of unobserved starting points, we refer the interested readers to the Appendix.

[8] A binary relation $\mathcal{N}$ is reflexive if $(x, x) \in \mathcal{N}$ for all $x \in X$. It is symmetric if $(y, x) \in \mathcal{N}$ whenever $(x, y) \in \mathcal{N}$.

[9] Symmetry does not necessarily imply that the exogenous network is symmetric. In fact, symmetry may be the result of consumers' symmetric associative memory. That is, if a consumer is recommended y when she considers x, she may recall x when she considers y even in the absence of a direct recommendation. As discussed in the introduction, the availability of y may be necessary for recollection. Recently, Kahana (2002) finds evidence suggesting that associative memory is indeed symmetric.

personalized recommendations or the use of different stores. For example, while some consumers use Netflix, others use Hulu for movies. Second, even if the exogenous network is the same for everyone (e.g., everyone uses Netflix and gets the same recommendations), it is still possible that different consumers pay attention to different recommended alternatives.

Suppose the consumer who is faced with the choice problem (S, x) considers all reachable alternatives from x in S in her perceived network. If the perceived network is the same as the exogenous network, this will correspond to a consumer who is not internally constrained (i.e., no search or attention cost), but externally constrained by the network structure. The external constraint imposed by the product network implies that no matter how much search the consumer does, there may be some alternatives she will never discover. The consumer's consideration set can be written as

$$N_x(S) = \{y \in S | \, \exists \, \{x_1, ..., x_k\} \subseteq S \text{ s.t. } x_1 = x, x_k = y, \text{ and } (x_i, x_{i+1}) \in \mathcal{N} \text{ for } i < k\}. \quad (3.1)$$

The set $N_x(S)$ denotes all the reachable alternatives from x in S. That is, if $y \in N_x(S)$, then there exists a sequence of linked alternatives connecting x to y (i.e., a path from x to y) in S. We can think of $N_x(S)$ as an endogenously determined consideration set of a consumer who is faced with the choice problem (S, x). We say that c is a network choice if there is a strict preference relation $\succ$ and a product network $\mathcal{N}$ such that for any choice problem (S, x), $c(S, x)$ maximizes $\succ$ in $N_x(S)$.

Definition 3.2.1. *A choice function c is a **network choice** if there exist a strict preference $\succ$ and a reflexive, symmetric binary relation $\mathcal{N}$ on X and such that*

$$c(S, x) = \text{argmax}(\succ, N_x(S))$$

where $N_x(S)$ is defined as in (3.1).

Notice that if a product network is complete, that is $(x, y) \in \mathcal{N}$ for all $x, y \in X$, then $N_x(S) =$

S for all S, and hence the induced choice behavior is rational. Therefore, the standard model is a special case of network choice. When the product network is incomplete, seemingly anomalous choice behaviors could be observed. The following example illustrates network choice and the type of choice behavior it allows.

Example 3.2.1. *Altan, who is a six-year-old boy, faces a linear product network which is described in Figure 3.1.*[10]

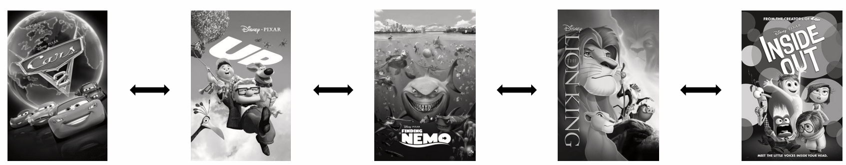

Figure 3.1: Altan's product network

Altan's most favorite movie is Inside Out *and his second favorite* Cars.[11] *Altan's starting point is the last movie his older brother has seen. If all the movies are currently available in Netflix, Altan will be aware of all the movies through Netflix's recommendation (because the entire network is connected). Hence, he chooses his most favorite movie* (Inside Out) *independent of the starting point. If* Nemo *becomes unavailable in Netflix, the implied network is no longer connected, and hence the starting point matters. In this case, if the last movie his older brother has seen is* Lion King*, then* Inside Out *is still reachable and will be chosen. However, if the starting point is* Up *or* Cars*, Altan ends up watching* Cars*.*

In the above example, Altan exhibits two types of choice patterns. First, changing the starting point affects his final choice (when the choice set is fixed):

$$\text{Inside Out} = c(S, \text{Lion King}) \neq c(S, \text{Up}) = \text{Cars}.$$

[10]We imagine that Netflix recommends at most two available movies in this example. If the movie is one of the extreme points of the network, Netflix provides only one recommendation.

[11]His entire preference is Inside Out $\succ$ Cars $\succ$ Nemo $\succ$ Up $\succ$ Lion King.

where $S = \{\text{Cars, Up, Lion King, Inside Out}\}$. Second, including irrelevant alternatives affects his choice behavior (when the starting point is fixed):

$$\text{Inside Out} = c(S \cup \text{Nemo}, \text{Up}) \neq c(S, \text{Up}) = \text{Cars}$$

These two choice patterns will play a crucial role in our analysis later on.

3.2.1 Characterization of Search Behavior

Before we provide behavioral postulates to characterize our choice procedure, we would like to investigate the properties of consideration sets $(N_x(S))$ in this model. These properties will help us to identify and motivate behavioral postulates in the next subsection.

We first define the notion of starting-point-dependent consideration sets. Let $\mathcal{X}(y)$ be the set of all subsets of X including y. We say that $\{\Gamma_x\}_{x \in X}$ is a starting-point-dependent consideration set mapping if for any $x \in X$, the mapping $\Gamma_x : \mathcal{X}(x) \to \mathcal{X}(x)$ satisfies $\Gamma_x(S) \subseteq S$ for all $S \in \mathcal{X}(x)$. Now the question is what type of properties on the consideration sets guarantee that they are induced by a search over a product network.

To provide an answer for this question, we introduce three properties. The first property says that the set of all reachable alternatives does not shrink as the choice problem expands. That is, $\Gamma_x(T) \subseteq \Gamma_x(S)$ whenever $T \subseteq S$ for a fixed starting point. The second property says that if y belongs to $\Gamma_x(S)$, the consideration set $\Gamma_y(S)$ must be equal to $\Gamma_x(S)$. In our model, if y is reachable from x in S, then any alternative that is reachable from y must also be reachable from x and vice versa. The last property says that if z belongs to $\Gamma_x(S)$ but not to $\Gamma_x(S \setminus y)$, then y must be considered even in the absence of z, i.e., $y \in \Gamma_x(S \setminus z)$. In addition, the property says that z must be considered in the absence of x when y is the starting point, i.e., $z \in \Gamma_y(S \setminus x)$. In our model, the "if part" of the property implies that y must be on the path from x to z in the product network. Hence, z must be reachable from y and y must be reachable from x.

We now formally state these properties and provide the result.

A.1 (Upward Monotonicity) $\Gamma_x(T) \subseteq \Gamma_x(S)$ whenever $T \subseteq S$.

A.2 (Symmetry) If $y \in \Gamma_x(S)$, then $\Gamma_y(S) = \Gamma_x(S)$.

A.3 (Path Connectedness) If $z \in \Gamma_x(S)$ and $z \notin \Gamma_x(S \setminus y)$, then $y \in \Gamma_x(S \setminus z)$ and $z \in \Gamma_y(S \setminus x)$ provided that x, y, and z are distinct alternatives.

Lemma 3.2.1. $\{\Gamma_x\}_{x \in X}$ *satisfies Upward Monotonicity, Symmetry, and Path Connectedness if and only if there exists a reflexive, symmetric binary relation* $\mathcal{N}$ *such that* $\Gamma_x(S) = N_x(S)$ *for all* $S \in \mathcal{X}$ *and* $x \in S$.[12]

Notice that these properties are defined on consideration sets which are usually unobservable. Nevertheless, Lemma 3.2.1 can be used to determine whether a consumer follows network choice when we have data on the consumer's consideration sets. For example, Reutskaja et al. (2011) find that the average number of items seen by the subjects increases with set size. This suggests an evidence for Upward Monotonicity. In addition, Lemma 3.2.1 also allows us to compare our model with existing limited consideration models. We discuss the relationship with other models in Section 3.3.

3.2.2 Characterization of Choice Behavior

As we discussed before, it is unlikely that an outside observer will have much information on implied consideration sets. We now propose three simple axioms on observed choice behavior. We then discuss how each axiom is related to three properties on consideration sets discussed above.

The first axiom is similar to standard contraction axiom (also known as α-axiom; Sen, 1971) for a fixed starting point. It says that if the starting point is chosen in some choice problem, then it must also be chosen in any subset of it as long as it stays as the starting point.

Axiom 3.2.1. *(Starting Point Contraction) If* $c(S, x) = x$ *and* $T \subseteq S$*, then* $c(T, x) = x$.

[12] All proofs are provided in the Appendix.

Axiom 3.2.1 is much weaker than the α-axiom. This axiom is silent when an alternative different from the starting point is chosen. On the other hand, α-axiom requires that removing irrelevant alternatives from a menu cannot affect the choice irrespective of which alternative was chosen. But in our network model, this conclusion is no longer true when the chosen alternative is different from the starting point since an alternative that is considered in a bigger choice set is not necessarily considered in a smaller choice set.

Axiom 3.2.1 is directly implied by the first property of consideration sets (A.1). To see this, suppose x is selected when the choice problem is (S, x). This means the starting point is the best among all reachable alternatives. If some alternatives are removed, the number of possible paths decreases, and hence the consideration set can only get smaller. Since the starting point is always available, and it was chosen when the choice set was larger, it must also be chosen when the choice set is smaller.

The second axiom says that if y is chosen in some choice problem (T, x), then the choices corresponding to choice problems (S, x) and (S, y) must be the same for all S containing T. In other words, once the starting point is abandoned for some alternative, replacing the original starting point with the chosen alternative should not affect choices in larger choice sets.

Axiom 3.2.2. *(Replacement) If $c(T, x) = y$ and $T \subseteq S$, then $c(S, x) = c(S, y)$.*

To see why Axiom 3.2.2 holds in our model, notice that the first and second properties of consideration sets imply a property which we call strong symmetry: if $y \in \Gamma_x(T)$ and $T \subseteq S$, then $\Gamma_x(S) = \Gamma_y(S)$. That is because if $y \in \Gamma_x(T)$ and $T \subseteq S$, then $y \in \Gamma_x(S)$ by A.1, and hence $\Gamma_x(S) = \Gamma_y(S)$ by A.2. Now since y is chosen (and hence considered) when the choice problem is (T, x), by strong symmetry, for any $S \supseteq T$, the consideration sets corresponding to choice problems (S, x) and (S, y) are the same. Therefore, the consumer must choose the same alternative.

From Example 3.2.1, we can see that the network choice model allows for choice reversal patterns which are not allowed by the standard model. This also distinguishes our model from reference-dependent choice models of Tversky and Kahneman (1991) and Masatlioglu and Ok

(2005, 2014). Both of these models assume that given a reference point the decision maker maximizes a reference-dependent utility function, and hence choices satisfy the weak axiom of revealed preference once the reference point is fixed.

Suppose we observe that removing y from choice problem (S, x) leads to a choice reversal, i.e., $y \neq c(S, x) \neq c(S \setminus y, x)$. That is, the presence of y changes the choice even though y is not chosen. In this case, we say that y is *influential* in the choice problem (S, x). The next axiom provides a structure on observed choice reversals by stating that if y is influential in (S, x), then x cannot be influential in (S, y).

Axiom 3.2.3. *(Asymmetry of Influence) If y is influential in (S, x), then x is not influential in (S, y).*

This axiom, together with other axioms, imply that the alternatives in S can be partially ordered by their relative influence in choices. If y causes a choice reversal in (S, x), then y is thought to be more influential than x in S. Hence, the removal of x from (S, y) should not affect the choice. Notice that the partial order of alternatives may depend on the choice set, i.e., it is possible that y is more influential than x in S, but x is more influential than y in T. Formally, let $y \gg_S x$ if $y \neq c(S, x) \neq c(S \setminus y, x)$. Then Axiom 3.2.3 is equivalent to the asymmetry of this binary relation. We can show that all the axioms together imply that $\gg_S$ is transitive. Hence, $\gg_S$ is a strict partial order on S.[13]

Notice that Axiom 3.2.2 disciplines starting point dependence while Axiom 3.2.3 imposes a structure on the violations of the Weak Axiom of Revealed Preference (WARP) for different starting points. However, one should expect that these two types of deviations from rationality are linked to each other. To illustrate why this relation is meaningful in our model, consider a network in which no two alternatives are linked. The choice behavior induced by such a product network exhibits extreme starting point dependence, and no alternative can be influential in such a network.

The next axiom provides a formal link between the two types of deviations from rationality. It states that if y is influential in (S, x), then there must be at least one $T \subset S$ such that (T, x) and

[13]To see why Axiom 3.2.3 holds in our model, notice that if y is influential in (S, x), then it must be the case that $c(S, x) \in \Gamma_x(S)$ but $c(S, x) \notin \Gamma_x(S \setminus y)$. By A.1, A.2, and A.3, $\Gamma_x(S) = \Gamma_y(S)$, and hence $c(S, x) = c(S, y)$. Moreover, by A.1 and A.3, $c(S, y) \in \Gamma_y(S \setminus x) \subseteq \Gamma_y(S)$. Hence, $c(S, y) = c(S \setminus x, y)$ as desired.

(T, y) do not exhibit starting point dependence. In other words, a deviation from rationality on one dimension (a violation of WARP for a fixed starting point) precludes a deviation from rationality on another dimension (starting point dependence) for some choice problem. To introduce the axiom, we say that x and y are *distant* in S if for all proper subsets T of S the choices from (T, x) and (T, y) are distinct.

Axiom 3.2.4. *(Irrelevant if Distant) If x and y are distant in S, then y is not influential in (S, x).*

Intuitively, since distant alternatives x and y lead to different choices, they must be "far away" from each other in the product network. Therefore, either of these alternatives cannot be influential in this choice problem when the other alternative is the starting point.[14]

The following theorem provides a foundation for the network choice model.

Theorem 3.2.1. *A choice function c satisfies Starting Point Contraction, Replacement, Asymmetry of Influence, and Irrelevant if Distant if and only if it is a network choice.*

Theorem 3.2.1 shows that network choice is captured by four simple behavioral postulates. This makes it possible to test our model non-parametrically by using the standard revealed preference technique. Next, we derive the decision maker's preferences and network from observed choice data.

3.2.3 Revealed Preference and Network

In this section, we discuss how we can reveal preference and network from choice data given that the consumer follows the network choice. The standard theory suggests that choices directly reveal preferences. That is, x is preferred to y when x is chosen in the presence of y. To justify such an inference, one must implicitly assume that y is considered. In our model, the decision maker is constrained by the network. As a result, the decision maker might not compare all available alternatives before making a choice. Therefore, eliciting preference from choices is no longer

[14]Axiom 3.2.4 is an implication of our model as y is influential in (S, x) if and only if $c(S, x) \in \Gamma_x(S)$ but $c(S, x) \notin \Gamma_x(S \setminus y)$. Hence, by A.2 and A.3, $\Gamma_x(S \setminus c(S, x)) = \Gamma_y(S \setminus c(S, x))$ which implies that $c(T, x) = c(T, y)$ for $T = S \setminus c(S, x)$.

trivial. The next example illustrates that there may be multiple preferences representing given choice behavior.

Example 3.2.2. *Now consider another decision maker, Mehmet, who is very fond of Lego movies:* Lego Movie, Batman, Ninjago*. While we do not have any knowledge about his underlying choice procedure, we can observe his choice behavior. He always chooses* Ninjago *independent of the starting point as long as it is available. However, in the absence of* Ninjago*, his choice is dictated by the starting point. That is, he watches what he encounters first.*

Since Mehmet's behavior satisfies Starting Point Contraction, Replacement, Asymmetry of Influence, and Irrelevant if Distant, Theorem 3.2.1 guarantees that a network choice representation of his behavior exists. The fact that Mehmet chooses Ninjago from each binary choice problem implies that Ninjago is directly linked to both movies. On the other hand, since Mehmet chooses different alternatives from $\{$Lego Movie, Batman$\}$ depending on the starting point, there is no link between Lego Movie and Batman (notice there is a path between them through Ninjago). Hence, his choices reveal his entire network. This is a general feature of the baseline model. To reveal the underlying network, it suffices to check binary choice data with different starting points. If the decision maker's choices between two options are independent of starting points, then we reveal that these two options are directly linked. Otherwise, there is no direct link (see Figure 3.2).

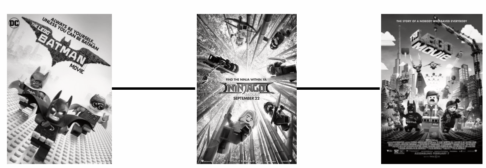

Figure 3.2: Mehmet's product network (Example 3.2.2)

While the entire network is pinned down, revealed preference is not unique in this example. Indeed, Mehmet's preference may be either of the following: Ninjago $\succ_1$ Lego Movie $\succ_1$ Batman and Ninjago $\succ_2$ Batman $\succ_2$ Lego Movie. Since there can be multiple preferences representing choice behavior, we need to define what we mean by revealed preference. Following Masatlioglu et al. (2012), we say that an option is revealed preferred to another option if the first option is ranked higher than the second one in all possible representations.

Definition 3.2.2. *Suppose c is a network choice and let $\{(\succ_i, \mathcal{N})|\ i = 1, \ldots, N\}$ be all possible representations of c. Then, we say that x is revealed to be preferred to y if $x \succ_i y$ for all $i = 1, \ldots, N$.*

Once we reveal the network, we can compare two options in terms of preference only if there is a direct link or a path between them. If there is no link or path between these two options, there will be no choice problem in which these alternatives are considered at the same time, and hence we cannot tell which option is more preferred.

One might think that if there is a path between two alternatives, then we can reveal which one is more preferred. However, Example 3.2.2 illustrates that the existence of a path is not enough either. In that example, there is a path between Lego Movie and Batman, but we still cannot tell which alternative is more preferred. We can reveal preference between two alternatives if and only if there is a choice set in which (i) a direct link or a path between the two alternatives exists, and (ii) one alternative is chosen over the other.

Consider an observation where the decision maker chooses x when y is the starting point, i.e., $x = c(S, y)$. This observation satisfies both conditions: (i) there is a path between x and y in S, and (ii) x is chosen over y. We can list many other choice patterns that satisfy both (i) and (ii), and hence reveal preference. Consider the following sets of observations.

1. $x = c(S, z)$ and $y = c(T, z)$ where $T \subsetneq S$,

2. $x = c(S, z)$ and $c(T, y) = c(T, z)$ where $T \subseteq S$,

3. $x = c(S, z)$ and $y \neq c(T, z) \neq c(T \setminus y, z)$ where $T \subseteq S$.

In all the observations above we learn that there is a path between y and z in T. This guarantees that y must be considered in any choice set containing T when z is the starting point. Hence, the observation that x is chosen in the choice problem (S, z) implies that x is preferred to y. The list above is far from being exhaustive. For example, suppose $x = c(S,t)$, $c(T,z) = c(T,y)$, and $c(T,z) = c(T',t)$ where $T, T' \subseteq S$. These observations tell us that there is a path from both t to z and z to y in S. Hence, y is considered when x is chosen.

It may appear that revealed preference can be quite complicated in this setup. However, if we have access to complete choice data and Axioms 3.2.1-3.2.4 are satisfied, then by applying the axioms we can verify that all the above observations are possible only if there exists a choice set such that x is chosen when y is the starting point. As an example, Replacement guarantees that $c(S,y) = x$ in the case of Observation 1. The next proposition states that with complete choice data, revealed preference can be summarized by one simple observation: x is preferred to y if and only if x is chosen when y is the starting point.

Proposition 3.2.1. *Suppose c is a network choice. Then,*

- *x is revealed preferred to y if and only if $c(S,y) = x$ for some S,*
- *x is revealed to be linked to y if and only if $c(\{x,y\},x) = c(\{x,y\},y)$.*

Proposition 3.2.1 provides a necessary and sufficient condition for revealed preference in our model. As discussed, the result assumes that the analyst observes entire choice behavior. With a limited dataset, we can use the observations listed above to reveal consumer preference.

We now revisit Example 3.2.1 to illustrate how to utilize Proposition 3.2.1. We pretend that we do not know either Altan's preferences nor his network structure and try to reveal them from his observed choices. Since Altan chooses Inside out when everything is available and the starting point is Cars, this implies that Altan prefers Inside Out over Cars. Similarly, since Cars is chosen when Nemo is the starting point and the menu is $\{\text{Cars}, \text{Nemo}, \text{Up}\}$, Cars is better than Nemo. Choosing Nemo from the choice problem $(\{\text{Nemo}, \text{Up}\}, \text{Up})$ reveals that Nemo is preferred to Up. Finally, $c(\{\text{Nemo}, \text{Up}, \text{Lion King}\}, \text{Lion King}) = \text{Up}$ yields that Lion King is the worst one. Hence, we infer Altan's entire preferences. It is routine to check that all the binary choice problems reveal the product network given in Figure 3.1.

3.3 Limited Search

Netflix offers thousands of movies which makes it impossible to do an exhaustive search. When people are confronted with an overwhelming number of options, they must implement a limited search. In this section, we consider an agent who terminates her search after a certain number of rounds due to limitations such as time pressure or limited cognitive capacity.

We assume that the consumer starts searching from a certain starting point and considers all the alternatives that are linked to the starting point. The decision maker stops search after K steps where $K \geq 2$. For example, if $K = 2$, then the decision maker only considers the starting point and the alternatives which are directly linked to the starting point. If the number of steps is larger than the number of alternatives, then this model reduces to our baseline model. We provide a characterization for K-step network choice where K is fixed and discuss the properties of the consideration sets induced by K-step search.

Consider a consumer faced with the choice problem (S, x). The K-step consideration set is given by

$$\begin{aligned} N_x^K(S) = \{y \in S | \exists \{x_1, \dots, x_k\} \subseteq S \text{ such that } x_1 = x, x_k = y, k \leq K, \\ \text{and } (x_i, x_{i+1}) \in \mathcal{N} \text{ for } i < k\} \end{aligned} \tag{3.2}$$

We say that a consumer makes a K-step network choice if the consumer picks the best element in the K-step consideration set.

Definition 3.3.1. *A choice function c is a K-step network choice if there exist a strict preference $\succ$ and a reflexive, symmetric binary relation $\mathcal{N}$ on X such that*

$$c(S, x) = \text{argmax}(\succ, N_x^K(S))$$

where $N_x^K(S)$ is as in (3.2).[15]

[15]If $K = 2$, then it is without loss of generality to assume that $\mathcal{N}$ is symmetric.

The following example illustrates the properties of K-step network choice.

Example 3.3.1. *Efe is the little brother of Altan from Example 3.2.1. Both brothers share the same preferences and face the same product network. However, Efe always stops the search after two steps ($K = 2$) due to his cognitive limitations. Even though he does a limited search, his choice behavior satisfies WARP for each starting point. Hence, Efe acts "as if" he is a classical utility maximizer for a fixed starting point.*

This example highlights the power of our choice data. The richness of our data prohibits us making wrong claims about the revealed preference. For example, assume that we only observe Efe's choice data when the starting point is Nemo. In this case, Efe's behavior satisfies WARP. If we apply the standard revealed preference, we must conclude that Nemo is strictly preferred to both Inside Out and Cars, which are the best two options for Efe. In other words, we will mistakenly reveal that Inside Out is worse than Nemo even though it is the best alternative for Efe. With a richer dataset that includes observations with multiple starting points, we can observe that Efe's choices depend on the starting point in contrast to the standard model.

3.3.1 Characterization of Search Behavior

Before moving on to the characterization of choice behavior, we first show the properties the consideration sets arising from K-step network choice satisfy. First, notice that Upward Monotonicity (A.1) and Path Connectedness (A.3) must still be satisfied in the modified model. However, the original Symmetry property (A.2) is no longer satisfied. That is because in K-step search model, even if y is reachable from x in K steps and z is reachable from y in K steps, this does not necessarily imply that z is reachable from x in K steps. If the choice set has fewer than K alternatives, then Symmetry follows.

For any set S, we use the notation $\mathcal{P}_{\leq K}(S)$ to denote all nonempty subsets of S with at most K elements.

B.2 (K-Symmetry) If $y \in \Gamma_x(S)$ and $S \in \mathcal{P}_{\leq K}(X)$, then $\Gamma_y(S) = \Gamma_x(S)$.

Finally, consideration sets in K-step search model also satisfy K-Reduction which has no analog in the baseline model. This property says that if y is reachable from x in S, then there exists $T \subseteq S$ consisting of at most K alternatives such that y is reachable from x in T. To see why this must be true, consider the set consisting of x, y, and the alternatives connecting x to y. This set must have at most K alternatives.

B.4 (K-Reduction) If $y \in \Gamma_x(S)$, then there exists $T \in \mathcal{P}_{\leq K}(S)$ such that $y \in \Gamma_x(T)$.

Lemma 3.3.1 shows that if a collection of consideration set mappings $\{\Gamma_x\}_{x \in X}$ satisfies the four properties described above, then we can treat them as a K-step consideration set on a product network.

Lemma 3.3.1. $\{\Gamma_x\}_{x \in X}$ *satisfies Upward Monotonicity, K-Symmetry, Path Connectedness, and K-Reduction if and only if there exists a reflexive, symmetric binary relation $\mathcal{N}$ such that* $\Gamma_x(S) = N_x^K(S)$ *for all* $S \in \mathcal{X}$ *and* $x \in S$.

Lemma 3.3.1 is useful in comparing our model with existing models of limited consideration. In the recent literature on limited consideration, a decision maker chooses the best alternative from a small subset of available alternatives. Such models include rational shortlisting (Manzini and Mariotti, 2007), considering alternatives that belong to the best category (Manzini and Mariotti, 2012), considering alternatives that are optimal according to some rationalizing criteria (Cherepanov et al., 2013), and limited attention (Masatlioglu et al., 2012; Lleras et al., 2017). While all of these models have an element of limited attention, choices are not affected by a starting point in these models. Even though the domains of these models are different from ours, we contrast our model with these models by fixing the starting point.

These models satisfy one of the two following properties. The first condition says that the consideration set is unaffected by the removal of an alternative which does not attract attention.

$$x \notin \Gamma(S) \text{ implies } \Gamma(S) = \Gamma(S \setminus x).$$

The second property captures the idea that attention is relatively scarcer in larger choice sets. That is, if an alternative attracts attention in a larger set, it also attracts attention in subsets of it in which it is included.

$$x \in \Gamma(S) \text{ implies } x \in \Gamma(T) \text{ if } x \in T \subseteq S.$$

It is routine to show that for a fixed starting point, Upward Monotonicity and Path Connectedness imply the first condition. On the other hand, our model assumes the opposite of the second condition.

Tyson (2013) considers a model where each menu has an associated preference relation. The smaller the choice sets the more fine-grained the preferences are. The intuition is that in more complex decision problems it is more difficult to compare alternatives. In the first stage, menu-dependent preferences are used to eliminate some alternatives. In the second stage, ties are broken by a salience relation. From the four properties that are satisfied by consideration sets in our model, only Upward Monotonicity holds in his model.

3.3.2 Characterization of Choice Behavior

We now propose axioms on choices characterizing K-step network choice for a fixed K. Even in the presence of bounded rationality, three of our original four axioms (Starting Point Contraction, Asymmetry of Influence, and Irrelevant if Distant) hold in K-step network choice model for arbitrary K. On the other hand, Replacement holds if the choice set does not have more than K alternatives.

The fact that Starting Point Contraction still holds must be obvious as it is an implication of Upward Monotonicity, which is still true for K-step consideration sets. To see why Asymmetry of Influence is still true, suppose y is influential in (S, x) and x is influential in (S, y), i.e., $y \neq c(S, x) \neq c(S \setminus y, x)$ and $x \neq c(S, y) \neq c(S \setminus x, y)$. This is possible if and only if $c(S, x) \in \Gamma_x(S) \setminus \Gamma_x(S \setminus y)$ and $c(S, y) \in \Gamma_y(S) \setminus \Gamma_y(S \setminus x)$. By A.3, $c(S, x) \in \Gamma_y(S \setminus x)$ and $c(S, y) \in \Gamma_x(S \setminus y)$. Since $c(S, x) \notin \Gamma_x(S \setminus y)$, $c(S, x) \neq c(S, y)$ follows. Now by A.1, $c(S, y) \in \Gamma_x(S)$ and $c(S, x) \in \Gamma_y(S)$. But then the former implies $c(S, x) \succ c(S, y)$ and the latter implies $c(S, y) \succ c(S, x)$, a contradiction.

Finally, we show that Irrelevant if Distant is still true in the K-step network choice model. Suppose y is influential in S. As before, this implies that $c(S, x) \in \Gamma_x(S) \setminus \Gamma_x(S \setminus y)$. By A.3, $y \in \Gamma_x(S \setminus c(S, x))$. By B.4, there exists $T \subseteq S \setminus c(S, x)$ with $|T| \leq K$ such that $y \in \Gamma_x(T)$. Hence, by B.2, $\Gamma_x(T) = \Gamma_y(T)$ and $c(T, x) = c(T, y)$ follows. We conclude that x and y are not distant in S, as required by the axiom.

While Starting Point Contraction is still necessary for the model, we need to replace it with a stronger axiom. This stronger axiom states that every choice set has a dominant alternative: for any set S, there exists an alternative x^* that is the best in S. Suppose given a choice problem (T', z), the consumer picks x^*. Then, if we extend the choice set and consider the choice problem (T, z), the consumer must still consider x^*. Since x^* is the best element in S, if the consumer picks an alternative that belongs to S, it must be x^*.

Axiom 3.3.1. *(Dominant Alternative) For any S, there exists $x^* \in S$ such that for any $z \in T' \subseteq T$, if $c(T', z) = x^*$ and $c(T, z) \in S$, then $c(T, z) = x^*$.*

Axiom 3.3.1 is much stronger than Starting Point Contraction. On the other hand, Starting Point Contraction together with Replacement imply Axiom 3.3.1. Hence this axiom holds in our baseline model too. Since the original Replacement axiom no longer holds in K-step network choice model, Axiom 3.3.1 is no longer implied by the other axioms, and hence it must be stated independently.

The original Replacement axiom might not hold due to bounded rationality. Axiom 3.3.2 is a modification of Replacement. It says that if the starting point is abandoned for another alternative in some choice set T, then replacing the original starting point with the chosen alternative does not alter the choice for any choice set containing T as long as the choice set does not have more than K alternatives.

Axiom 3.3.2. *(K-Replacement) If $c(T, x) = y$ and $T \subseteq S \in \mathcal{P}_{\leq K}(X)$, then $c(S, x) = c(S, y)$.*

To see why it holds in our model, suppose the consumer chooses y when the choice problem is

(T, x). Then, y is reachable from x in T in K steps. By Upward Monotonicity, if $S \supseteq T$, then y is considered in the choice problem (S, x). By K-Symmetry, if S has fewer than K alternatives, the consideration sets corresponding to choice problems (S, x) and (S, y) must be the same. Therefore, the consumer must make the same choice in both choice problems.

We now introduce a new axiom, K-Reduction, which has no analog in the baseline model. Hence one can think of this new axiom as a way of capturing bounded rationality in this model. It states that if y is chosen when the choice problem is (S, x), then there exists $T \subseteq S$ with at most K alternatives such that y is chosen when the choice problem is (T, x). This is an easy consequence of the K-step network model as the chosen alternative must be reachable in at most K steps.

Axiom 3.3.3. *(K-Reduction) If $c(S, x) = y$, then there exists $T \in \mathcal{P}_{\leq K}(S)$ such that $c(T, x) = y$.*

Theorem 3.3.1 provides a foundation for K-step network choice.

Theorem 3.3.1. *A choice function c satisfies Dominant Alternative, Asymmetry of Influence, Irrelevant if Distant, K-Replacement, and K-Reduction if and only if it is a K-step network choice.*

Since the only two axioms which depend on K are K-Replacement and K-Reduction, our characterization allows us to decompose the effects of an increase in K (or more search) into two components. Notice that as K increases, K-Replacement becomes stronger while K-Reduction becomes weaker. Since K-Replacement imposes a structure on choices across different starting points for a fixed choice set, while K-Reduction connects choices across different choice sets for a fixed starting point, the choices of the consumer with a higher K will exhibit less starting point dependence and more violations of WARP for a fixed starting point.

Theorem 3.3.1 covers an important special case ($K = 2$) which has previously been studied by Masatlioglu and Ok (2014). However, unlike us they assume observability of choices when there is no status quo. This assumption seems implausible in our domain as every search has a starting point. The characterization of their model without this assumption was an open question to which Theorem 3.3.1 provides an answer. Notice that when $K = 2$, K-Replacement simply states that if $c(\{x, y\}, x) = y$, then $c(\{x, y\}, y) = y$. This is implied by the Dominant Alternative axiom if we set $y = z$, $T' = \{y\}$, and $T = S = \{x, y\}$. Moreover, 2-step network choice model does not exhibit choice reversals, and hence Asymmetry of Influence and Irrelevant if Distant are vacuously true. In addition, 2-Reduction can be written as "$c(S, x) = y \Rightarrow c(\{x, y\}, x) = y$." This leads us to the following corollary.

Corollary 3.3.1. *A choice function c satisfies Dominant Alternative and 2-Reduction if and only if it is a 2-step network choice.*

It is illustrative to compare the characterization in Masatlioglu and Ok (2014) with ours. Since they assume observability of choices with no starting points, their axiomatic structure looks quite different. The two axioms in their setup that can be translated to our setting are Weak Axiom of Revealed Preference (WARP) and Weak Status Quo Bias (WSQB). WARP states that if $c(S, x) = y$ and $\{x, y\} \subseteq T \subseteq S$, then $c(T, x) = y$. Obviously, WARP implies 2-Reduction but not vice versa. However, WARP and Dominant Alternative are independent axioms. Since WARP relates choices across different sets for a single starting point and Dominant Alternative relates choices across different sets and different starting points, it is easy to construct choice behavior that satisfies WARP but not Dominant Alternative. On the other hand, by Theorem 3.3.1, 3-step network choice satisfies Dominant Alternative, but it does not necessarily satisfy WARP. By Corollary 3.3.1, Dominant Alternative combined with 2-Reduction implies WARP.

In our setting, Weak Status Quo Bias states that "$c(\{x, y\}, x) = y \Rightarrow c(\{x, y\}, y) = y$." This is just 2-Replacement axiom and it is implied by Dominant Alternative as discussed above. Masatlioglu and Ok (2014) also discuss a stronger axiom which they call Strong Status Quo Bias (SSQB). Even though their model does not satisfy SSQB, replacing WSQB with SSQB in their axiomatic structure provides a characterization of Masatlioglu and Ok (2005). In our setting, SSQB states that if $c(S, x) = y$, then $c(S, y) = y$. This is an implication of Replacement which is satisfied by our baseline model. Recall that unlike Masatlioglu and Ok (2005) our baseline model does not necessarily satisfy WARP.

K-step model is also related to the work Tversky and Kahneman (1991). They introduce a reference-dependent choice model in which the reference point affects the utility of an individual by overweighting losses relative to the reference point. In contrast, the DM in our model has a fixed preference and the reference point affects the decision through the attention channel. In terms of choice behavior, their model allows cycles in choices across reference points, i.e., $c(\{x, y\}, x) = y$, $c(\{y, z\}, y) = z$, and $c(\{x, z\}, z) = x$. This type of choice behavior is not allowed in our model. For a fixed starting point, while our model allows choice reversals, their model satisfies WARP.

Dean et al. (2017) propose a reference-dependent model with limited attention. For a choice problem (S, x), the consideration set is given by $\Gamma_x(S) = (\Gamma(S) \cup x) \cap Q(x)$ where $x \in \Gamma(S)$ implies $x \in \Gamma(T)$ if $x \in T \subseteq S$. The consideration sets of this model violate all our properties. Indeed, they assume downward monotonicity which is the exact opposite of upward monotonicity property in our model. As opposed to our model, they allow the decision maker not to choose the reference point in a smaller choice set while choosing it in the larger choice set.

3.3.3 Revealed Preference and Network

In this section, we discuss how one can reveal preference and network given that the consumer follows K-step network choice. First, notice that since $K \geq 2$ the revelation of the network is exactly the same as in the baseline model for all K. In particular, x is revealed to be linked to y if and only if $c(\{x, y\}, x) = c(\{x, y\}, y)$, and x is revealed not to be linked to y if and only if $c(\{x, y\}, x) \neq c(\{x, y\}, y)$.

Recall Efe from Example 3.3.1. We again pretend that we only observe Efe's choices and try to reveal his preferences from his choices. We illustrate how we can infer his preference. Recall that

$$\begin{aligned} \text{c(\{Cars,Nemo,Up\},Up)} &= \text{Cars} \\ \text{c(\{Nemo,Up\},Up)} &= \text{Nemo} \end{aligned}$$

These choices immediately reveal that Nemo and Cars are strictly better than Up. Since both Nemo and Cars are linked to Up, Efe must have considered Nemo when he has chosen Cars in the first choice observation. Therefore, Cars must be revealed to be preferred to Nemo.

We now generalize this observation. For any starting point, an alternative that is chosen in a bigger set must be more preferred. That is an implication of Upward Monotonicity property of consideration sets. For any $x \neq y$, we define

$$xPy \text{ if there exists } z \in X \text{ and } T \subset S \subseteq X \text{ such that } c(S,z) = x \text{ and } c(T,z) = y$$

Let P_R denote the transitive closure of P. It is easy to see that if xP_Ry, then x must be revealed preferred to y. Proposition 3.3.1 says that if x is revealed to be preferred to y, then we must also have xP_Ry. In other words, Proposition 3.3.1 provides a characterization of the revealed preference in K-step search model.

Proposition 3.3.1. *(Revealed Preference) Suppose c is a K-step network choice. Then x is revealed to be preferred to y if and only if xP_Ry.*[16]

Proposition 3.3.1 highlights an interesting feature of K-step search model. The revealed preference of this model is independent of K. Even if we are unsure about the exact value of K for a decision maker, we can still do revealed preference analysis. This is useful in applications where we might have limited knowledge about K.

Since the baseline model is a special case of the K-step model where K is big enough, one might wonder the relationship between Proposition 3.2.1 and 3.3.1. In general, the revealed preference defined in Proposition 3.3.1 is richer than the one defined in Proposition 3.2.1. However, they coincide when K is big enough and we have complete choice data.

3.4 Conclusion

Many real life decision-making problems involve a search over a product network. In this paper, we show how one can reveal preference and network from individual choice data and provide characterizations of the models of decision making within a product network. We explore the cases of "perfectly rational" and "boundedly rational" consumers.

[16]The proof of Proposition 3.3.1 directly follows the proof of Theorem 3.3.1.

There are several interesting open questions. First, this paper only discusses symmetric links or undirected product networks. An obvious open question is how the implications of such a model change when the links are asymmetric. Second, while we treat the number of steps as an exogenously given, one can endogenize the number of search steps that the decision maker takes. It is plausible that the number of search steps depends on the complexity of a product network. Third, one can also think about alternative ways of modeling bounded rationality. For example, there may be a temptation ranking that determines which advertised products the decision maker considers.

Another avenue to explore is to study network choice with a random network. The randomness of a network could arise from two factors: (i) the exogenous product network that we take as given may be random (for example, Netflix's recommendation algorithm may produce random links between alternatives), (ii) the decision maker may pay random attention to presented alternatives.

3.5 Appendix A: Proofs

Proof of Lemma 3.2.1

Proof. ($\Leftarrow$) A.1: Suppose $y \in \Gamma_x(T)$. Then, there exists $\{x_1, \ldots, x_k\} \subseteq T \subseteq S$ such that $x_1 = x, x_k = y$, and $(x_i, x_{i+1}) \in \mathcal{N}$ for $i < k$. By definition, $y \in \Gamma_x(S)$.

A.2: Suppose $y \in \Gamma_x(S)$. This implies that there exists $\{x_1, \ldots x_j\} \subseteq S$ with $x_1 = x$, $x_j = y$, and $(x_i, x_{i+1}) \in \mathcal{N}$ for $i < j$. If $z \in \Gamma_y(S)$, then there exists $\{x_j, \ldots, x_k\} \subseteq S$ such that $x_j = y$, $x_k = z$, and $(x_i, x_{i+1}) \in \mathcal{N}$ for $j \leq i < k$. Consider $\{x_1, \ldots, x_j, \ldots, x_k\} \subseteq S$. It satisfies the conditions that $x_1 = x$, $x_k = z$, $(x_i, x_{i+1}) \in \mathcal{N}$ for $i < k$. Therefore, $z \in \Gamma_x(S)$. Now suppose $z \in \Gamma_x(S)$. First, let $x'_i = x_{j-i+1}$ for $i \leq j$. Then, $\{x'_1, \ldots, x'_j\} \subseteq S$ satisfies the conditions that $x'_1 = y$, $x'_j = x$, and $(x'_i, x'_{i+1}) \in \mathcal{N}$ for $i < j$. Since $z \in \Gamma_x(S)$, there exists $\{x'_j, \ldots, x'_k\} \subseteq S$ with $x'_j = x$, $x'_k = z$, and $(x'_i, x'_{i+1}) \in \mathcal{N}$ for $j \leq i < k$. Consider $\{x'_1, \ldots, x'_j, \ldots, x'_k\}$. It satisfies the conditions that $x'_1 = y$, $x'_k = z$, $(x'_i, x'_{i+1}) \in \mathcal{N}$ for $i < k$. Therefore, $z \in \Gamma_y(S)$.

A.3: Suppose $z \in \Gamma_x(S)$ and $z \notin \Gamma_x(S \setminus y)$. Since $z \in \Gamma_x(S)$, there exists $\{x_1, \ldots, x_k\} \subseteq S$

with $x_1 = x,\ x_k = z$, and $(x_i, x_{i+1}) \in \mathcal{N}$ for $i < k$. Furthermore, since $z \notin \Gamma_x(S \setminus y)$ there exists $j \in \{2, \ldots, k-1\}$ such that $x_j = y$. Consider $\{x_1, \ldots, x_j\} \subseteq S \setminus z$. It satisfies the conditions that $x_1 = x,\ x_j = y$, and $(x_i, x_{i+1}) \in \mathcal{N}$ for $i < j$. Therefore, $y \in \Gamma_x(S \setminus z)$. Now consider $\{x_j, \ldots, x_k\} \subseteq S \setminus x$. It satisfies the conditions that $x_j = y,\ x_k = z$, and $(x_i, x_{i+1}) \in \mathcal{N}$ for $j \leq i < k$. Therefore, $z \in \Gamma_y(S \setminus x)$.

($\Rightarrow$) Suppose $\{\Gamma_x\}_{x \in X}$ satisfies A.1–A.3. Let $(x, y) \in \mathcal{N}$ iff $y \in \Gamma_x(\{x, y\})$. Note that if $y \in \Gamma_x(\{x, y\})$, by A.2, $\Gamma_y(\{x, y\}) = \Gamma_x(\{x, y\})$. Therefore, for any $x, y \in X$, $(x, y) \in \mathcal{N}$ iff $(y, x) \in \mathcal{N}$. Given $\mathcal{N}$, we define $N_x(S)$ as in (3.1). Note that, by the previous part, $N_x(S)$ defined as such satisfies A.1-A.3. First, we show that $N_x(S) \subseteq \Gamma_x(S)$. Suppose $y \in N_x(S)$. Then, there exists $\{x_1, \ldots, x_k\} \subseteq S$ such that $x_1 = x,\ x_k = y$, and $x_{i+1} \in \Gamma_{x_i}(\{x_i, x_{i+1}\})$ for $i < k$. Therefore, by A.1, $x_{i+1} \in \Gamma_{x_i}(S)$ for $i < k$. By A.3, $\Gamma_{x_{k-1}}(S) = \Gamma_{x_{k-2}}(S) = \cdots = \Gamma_{x_1}(S)$. Then, $y = x_k \in \Gamma_{x_{k-1}}(S) = \Gamma_{x_1}(S) = \Gamma_x(S)$.

Now we show that $\Gamma_x(S) \subseteq N_x(S)$. The proof is by induction. First, note that if $y \in \Gamma_x(\{x, y\})$, then $y \in N_x(\{x, y\})$ by definition. Now suppose for all S with $|S| < n$ we have that $y \in \Gamma_x(S) \Rightarrow y \in N_x(S)$. Pick S with $|S| = n$ and suppose $y \in \Gamma_x(S)$. If there exists $z \in S \setminus \{x, y\}$ such that $y \in \Gamma_x(S \setminus z)$, then since $|S \setminus z| < n$ we have that $y \in N_x(S \setminus z)$ and, by A.1, $y \in N_x(S)$. Now suppose for all $z \in S \setminus \{x, y\}$, $y \notin \Gamma_x(S \setminus z)$. Pick one such z. Then, by A.3, $z \in \Gamma_x(S \setminus y)$ and $y \in \Gamma_z(S \setminus x)$. By induction hypothesis, $z \in N_x(S \setminus y)$ and $y \in N_z(S \setminus x)$. By A.1, $z \in N_x(S)$ and $y \in N_z(S)$. By A.2, $N_x(S) = N_z(S)$, and hence $y \in N_x(S)$. $\square$

Proof of Theorem 3.2.1

Necessity is obvious from the discussion in the main text. We prove sufficiency. For any $x \neq y$, define

$$xPy \text{ if and only if } \exists\, S \supseteq \{x, y\} \text{ such that } c(S, z) = x \text{ for all } z \in S$$

Claim 3.5.1. *P is transitive.*

Proof. Suppose $xPyPz$. Then there exist $S \supseteq \{x,y\}$ and $T \supseteq \{y,z\}$ such that $c(S,s)=x$ for all $s \in S$ and $c(T,t)=y$ for all $t \in T$. By Axiom 3.2.2, $c(S\cup T,s)=c(S\cup T,x)$ for all $s \in S$ and $c(S\cup T,t)=c(S\cup T,y)$ for all $t\in T$. Since $y \in S$, this implies that $c(S\cup T,s)=c(S\cup T,t)$ for all $s \in S$ and $t \in T$. Now if $c(S\cup T,t)=t$ for some $t \in T\setminus y$, then, by Axiom 3.2.1, $c(T,t)=t$ which is a contradiction. Similarly, $c(S\cup T,s)=s$ is not possible for $s \in S\setminus x$. Hence, $c(S\cup T,s)=c(S\cup T,t)=x$ for all $s\in S$ and $t \in T$. By definition, xPz and we are done. $\square$

Now let $\succ$ be a completion of P. We define $\mathcal{N}$ as

$$(x,y)\in\mathcal{N} \text{ if and only if } c(\{x,y\},x)=c(\{x,y\},y)$$

Note that $\mathcal{N}$ is reflexive and symmetric. Define $N_x(S)$ as

$$N_x(S)=\{y\in S|\exists\{x_1,\ldots,x_k\}\subseteq S \text{ such that } x_1=x, x_k=y, \\ \text{and } (x_i,x_{i+1})\in\mathcal{N} \text{ for } i<k\}$$

Claim 3.5.2. $c(S,x)\in N_x(S)$.

Proof. First, let $S=\{x,y\}$. If $c(\{x,y\},x)=x$, then the result is trivial. If $c(\{x,y\},x)=y$, by Axiom 3.2.2, $c(\{x,y\},x)=c(\{x,y\},y)$, and hence $y\in N_x(\{x,y\})$. Now suppose the claim is true for all S with $|S|<n$. Let S with $|S|=n$ be given. If there exists $z\in S$ such that $c(S,x)=c(S\setminus z,x)$, then by induction hypothesis, $c(S,x)\in N_x(S\setminus z)$, and by A.1, $c(S,x)\in N_x(S)$ so we are done. Now suppose for all $z\in S\setminus x$, $c(S,x)\neq c(S\setminus z,x)$. Notice that by Axiom 3.2.1 we must have $x\neq c(S,x)$. Pick $z\in S\setminus x$ such that $z\neq c(S,x)$. This is possible since $|S|>2$. Hence, z is influential in (S,x). By Axiom 3.2.3, x is not influential in (S,z). Moreover, by Axiom 3.2.4, there exists $T\subsetneq S$ such that $c(T,x)=c(T,z)$. Hence, by Axiom 3.2.2, $c(S,x)=c(S,c(T,x))=c(S,z)$. Since $x\neq c(S,x)$ and x is not influential in (S,z), we conclude that $c(S\setminus x,z)=c(S,z)=c(S,x)$. Now, by induction hypothesis, $c(S,x)\in N_z(S\setminus x)$, and by

A.1, $c(S,x) \in N_z(S)$. Furthermore, since by Axiom 3.2.4, $c(T,x) = c(T,z)$ for some $T \subsetneq S$, by induction hypothesis, $c(T,x) \in N_x(T) \cap N_z(T)$. Then, by A.2, $N_x(T) = N_{c(T,x)}(T) = N_z(T)$ which implies $z \in N_x(T)$, and hence by A.1, $z \in N_x(S)$. Finally, by A.2, $N_x(S) = N_z(S)$ and thus $c(S,x) \in N_x(S)$. $\square$

Claim 3.5.3. *If* $y \in N_x(S)$*, then* $c(N_x(S), y) = c(S,x)$

Proof. Suppose $y \in N_x(S)$. Then, there exists $\{x_1, \ldots, x_k\} \subseteq N_x(S) \subseteq S$ with $x_1 = x$, $x_k = y$ such that $c(\{x_i, x_{i+1}\}, x_i) = c(\{x_i, x_{i+1}\}, x_{i+1})$ for $i < k$. By Axiom 3.2.2, $c(N_x(S), x_1) = c(N_x(S), x_2) = \cdots = c(N_x(S), x_k)$ which implies $c(N_x(S), x) = c(N_x(S), y)$. By Claim 3.5.2, $c(S,x) \in N_x(S)$, and therefore $c(N_x(S), c(S,x)) = c(N_x(S), x)$. Furthermore, by Axiom 3.2.2, $c(S,x) = c(S, c(S,x))$, and by Axiom 3.2.1, $c(S, c(S,x)) = c(N_x(S), c(S,x))$. Therefore, $c(N_x(S), y) = c(N_x(S), x) = c(N_x(S), c(S,x)) = c(S,x)$. $\square$

Claim 3.5.4. $c(S,x) = \operatorname{argmax}(\succ, N_x(S))$.

Proof. By Claim 3.5.2, $c(S,x) \in N_x(S)$. By Claim 3.5.3, $c(N_x(S), y) = c(S,x)$ for all $y \in N_x(S)$. Therefore, by definition of $\succ$, $c(S,x) \succ y$ for all $y \in N_x(S) \setminus c(S,x)$. This completes the proof of the theorem. $\square$

Proof of Proposition 3.2.1

Proof. Notice that if c is a network choice and $c(S,y) = x$, then $c(N_y(S), z) = x$ for all $z \in N_y(S)$. The rest of the proof directly follows the proof of Theorem 3.2.1. $\square$

Proof of Lemma 3.3.1

Proof. Necessity can be easily verified as in the proof of Lemma 3.2.1. We prove sufficiency.

Let $(x,y) \in \mathcal{N}$ iff $y \in \Gamma_x(\{x,y\})$. By B.2, $\mathcal{N}$ is symmetric. Given $\mathcal{N}$, define $N_x^K(S)$ as in (3.2). Note that $N_x^K(S)$ defined as such must satisfy A.1, A.3, B.2, and B.4. We first show that $N_x^K(S) \subseteq \Gamma_x(S)$. Let $y \in N_x^K(S)$. Then, there exists $\{x_1, \ldots, x_k\} \subseteq S$ with $x_1 = x$, $x_k = y$,

$k \leq K$ such that $x_{i+1} \in \Gamma_{x_i}(\{x_i, x_{i+1}\})$ for $i < k$. Let $T = \{x_1, \ldots, x_k\}$. By A.1, $x_{i+1} \in \Gamma_{x_i}(T)$ for $i < k$. Since $|T| \leq K$, by B.2, $\Gamma_{x_1}(T) = \Gamma_{x_2}(T) = \cdots = \Gamma_{x_k}(T)$. Therefore, $y \in \Gamma_x(T)$. Finally, $T \subseteq S$ and A.1 imply $y \in \Gamma_x(S)$.

We now show that $\Gamma_x(S) \subseteq N_x^k(S)$. The proof is by induction. First, if $S = \{x, y\}$, the claim is obvious. Suppose the claim is true for all S with $|S| < n$. Let S with $|S| = n$ be given and suppose $y \in \Gamma_x(S)$. By B.4, there exists $T \subseteq S$ with $|T| \leq K$ such that $y \in \Gamma_x(T)$. If $T \subsetneq S$, then by induction hypothesis, $y \in N_x^K(T)$ and by A.1, $y \in N_x^K(S)$. If $T = S$, then $|S| \leq K$. Since there exists no proper subset T of S with $y \in \Gamma_x(T)$ we have that $y \notin \Gamma_x(S \setminus z)$ for all $z \in S \setminus x$. Pick z distinct from x and y. Now we have that $|S| \leq K$, $y \in \Gamma_x(S)$, and $y \notin \Gamma_x(S \setminus z)$. By A.3, $z \in \Gamma_x(S \backslash y)$ and $y \in \Gamma_z(S \backslash x)$. Then, by induction hypothesis, $z \in N_x^K(S \backslash y)$ and $y \in N_z^K(S \backslash x)$. By A.1, $z \in N_x^K(S)$ and $y \in N_z^K(S)$. Since $|S| \leq K$, by B.2, $N_x^K(S) = N_z^K(S) = N_y^K(S)$ which implies $y \in N_x^K(S)$. □

Proof of Theorem 3.3.1

Necessity is obvious from the discussion in the main text. We prove sufficiency. Let xPy if there exists $z \in T \subsetneq S$ such that $c(S, z) = x$ and $c(T, z) = y$.

Claim 3.5.5. *P is acyclic.*

Proof. Suppose $x_1 P x_2 P \cdots P x_n P x_1$. Then, there exists $\{T_i, T_i', z_i\}_{i=1}^n$ with $z_i \in T_i' \subsetneq T_i$ such that $c(T_i, z_i) = x_i$, $c(T_i', z_i) = x_{i+1}$ for $i < n$, and $c(T_n', z_n) = x_1$. Consider the set $S = \{x_1, \ldots, x_n\}$. For all $x \in S$, there exists $z \in T' \subsetneq T$ such that $c(T', z) = x$ and $c(T, z) \in S$, but $c(T, z) \neq x$. This contradicts Axiom 3.3.1. □

Let $\succ$ be a completion of P. Define $\mathcal{N}$ as

$$(x, y) \in \mathcal{N} \text{ if and only if } c(\{x, y\}, x) = c(\{x, y\}, y)$$

and $N_x^K(S)$ as

$$N_x^K(S) = \{y \in S | \, \exists \{x_1, \dots, x_k\} \subseteq S \text{ with } x_1 = x, x_k = y, k \leq K \\ \text{and } (x_i, x_{i+1}) \in \mathcal{N} \text{ for } i < k\}$$

Claim 3.5.6. $c(S, x) \in N_x^K(S)$.

Proof. First, note that if $c(\{x, y\}, x) = x$, then the claim is trivial. If $c(\{x, y\}, x) = y$, then by Axiom 3.3.2, $c(\{x, y\}, x) = c(\{x, y\}, y)$, which implies $y \in N_x^K(\{x, y\})$. Now suppose the claim is true for all S with $|S| < n$. Let S with $|S| = n$ be given. By Axiom 3.3.3, there exists $T \subseteq S$ with $|T| \leq K$ such that $c(T, x) = y$. If $T \subsetneq S$, then by induction hypothesis, $c(S, x) \in N_x^K(T)$, and by A.1, $c(S, x) \in N_x^K(S)$. Suppose $T = S$ so that $|S| \leq K$. Since there exists no proper subset T of S with $c(T, x) = y$, we must have that for all $z \in S \setminus x$, $c(S, x) = y \neq c(S \setminus z, x)$. Pick z distinct from x and y. Hence, z is influential in S. By Axiom 3.2.3, x is not influential in (S, z). Moreover, by Axiom 3.2.4, there exists $T' \subsetneq S$ such that $c(T', x) = c(T', z)$. Hence, by Axiom 3.3.2, $c(S, x) = c(S, c(T', x)) = c(S, z)$. Since $x \neq c(S, x)$ and x is not influential in (S, z), we conclude that $c(S \setminus x, z) = c(S, z) = c(S, x)$. Now, by induction hypothesis, $c(S, x) \in N_z^K(S \setminus x)$, and by A.1, $c(S, x) \in N_z^K(S)$. Furthermore, since by Axiom 3.2.4 $c(T', x) = c(T', z)$ for some $T' \subsetneq S$, by induction hypothesis, $c(T', x) \in N_x^K(T') \cap N_z^K(T')$. Then, by B.2, $N_x^K(T') = N_{c(T',x)}^K(T') = N_z^K(T')$ which implies $z \in N_x^K(T')$, and hence by A.1, $z \in N_x^K(S)$. Finally, by B.2, $N_x^K(S) = N_z^K(S)$ and thus $c(S, x) \in N_x^K(S)$. $\square$

Claim 3.5.7. *If $y \in N_x^K(S)$, then there exists $T \subseteq S$ such that $c(T, x) = c(T, y)$.*

Proof. Suppose $y \in N_x^K(S)$. Then, there exists $\{x_1, \dots, x_k\} \subseteq S$ with $x_1 = x$, $x_k = y$, $k \leq K$, and $(x_i, x_{i+1}) \in \mathcal{N}$ for $i < k$. By definition, $(x_i, x_{i+1}) \in \mathcal{N}$ if and only if $c(\{x_i, x_{i+1}\}, x_i) = c(\{x_i, x_{i+1}\}, x_{i+1})$. Let $T = \{x_1, \dots, x_k\}$. Since $c(\{x_i, x_{i+1}\}, x_i) = c(\{x_i, x_{i+1}\}, x_{i+1})$ we have that either $c(\{x_i, x_{i+1}\}, x_i) = x_{i+1}$ or $c(\{x_i, x_{i+1}\}, x_{i+1}) = x_i$. Then, since $|T| \leq K$, by Axiom 3.3.2, we have $c(T, x_1) = c(T, x_2) = \cdots = c(T, x_n)$. $\square$

Claim 3.5.8. $c(S,x) = \text{argmax}(\succ, N_x^K(S))$

Proof. By Claim 3.5.6, $c(S,x) \in N_x^K(S)$. Pick $y \in N_x^K(S)$. By Claim 3.5.7, there exists $T \subseteq S$ such that $c(T,x) = c(T,y)$. By definition of P, we have either $c(T,x) = y$ or $c(T,x)Py$. Furthermore, since $S \supseteq T$, either $c(S,x) = c(T,x)$ or $c(S,x)Pc(T,x)$. Since $\succ$ includes P, we have that either $c(S,x) = y$ or $c(S,x) \succ y$. This concludes the proof of the theorem. $\square$

3.6 Appendix B: Unobserved Starting Points

In the main text, we assume that we can observe the starting point of the consumer. Here, we investigate network choice with standard choice data. We first show that if we impose no structure on starting points, then any choice behavior can be justified. Suppose we observe choice function c where $c(S)$ is the element chosen by the consumer when the choice set is S. If our model is correct, then we must have $c(S) = c(S,x)$ where x is a starting point in S. If any alternative in S can be a starting point (i.e., there is no condition on how starting points in different sets are related), then we can let $c(S) = c(S, c(S))$ for all S. That is, the consumer always chooses the starting point. But then any choice behavior is allowed under this model. Hence, the model does not make any prediction.

In what follows, we impose a structure on starting points that helps us infer preferences and network with standard choice data. Following Salant and Rubinstein (2008) and Masatlioglu and Nakajima (2013), we assume that we observe induced choice correspondence where each possible choice corresponds to a different starting point. The induced choice correspondence reflects data available to an outside observer who knows that the choices of the decision maker are affected by the starting point, but lacks information about the actual starting point. Salant and Rubinstein explore a model in which the decision maker is allowed to make different choices under different frames. Given a choice correspondence C, the model is given by $C(S) = \{x \in S | \, x = c(S,f) \text{ for some } f \in F\}$ where F is the set of frames and $c(S,f)$ is frame dependent choice function. Masatlioglu and Nakajima use a similar idea with starting points.

Suppose the decision maker follows the network choice model denoted by c, but we do not observe her starting point. Let C stand for an induced choice correspondence where for every alternative x in $C(S)$, there exists a starting point y such that $x = c(S, y)$. In other words, x maximizes preference among all reachable alternatives from y in S.

Definition 3.6.1. *A choice correspondence C is an induced network choice if there exists a strict preference $\succ$ and a reflexive, symmetric binary relation $\mathcal{N}$ on X such that*

$$C(S) = \{x \in S | x = \text{argmax}(\succ, N_y(S)) \textit{ for some } y \in S\}$$

where $N_y(S)$ is defined as before.

Suppose we observe that the decision maker chooses different alternatives when faced with the same choice set. In standard theory, this would happen only if the decision maker is indifferent between chosen alternatives. However, in our model the decision maker with strict preference over all alternatives may still choose different alternatives when faced with the same choice set if the choice set is *not* connected (there are alternatives in the choice set such that no path between them exists).

Characterization

Before moving on to characterization, notice that using the symmetry property of consideration sets we can write the induced network choice as

$$C(S) = \{x \in S |\, x = \text{argmax}(\succ, N_x(S))\}$$

The alternative representation says that given a choice set S, an alternative x is chosen if and only if it is the best alternative among all the alternatives reachable from x in S. If y is reachable from x in S or vice versa, then the consideration sets corresponding to choice problems (S, x) and (S, y) are the same. Therefore, the original and the alternative representations are exactly the same.

We propose four simple axioms which characterize induced network choice. Axiom 3.6.1 is the standard contraction axiom. It says that if x is chosen when the choice set S, then x must also be chosen in any subset of S containing x.

Axiom 3.6.1. *(Contraction) If $x \in C(S)$, then $x \in C(T)$ for all $x \in T \subseteq S$.*

Note that Axiom 3.6.1 is a direct implication of the monotonicity property of consideration sets. Since consideration sets can only shrink as the choice set gets smaller, an alternative that is chosen in a bigger choice set must also be chosen in a smaller choice set as long as it is available.

Contraction axiom tells us what we should expect if x is chosen in some choice set S. Axiom 3.6.2 tells us what we should expect if x is not chosen. In particular, it posits the existence of an alternative y that dominates x. That is, if x is not chosen when the choice set is S, then there must exist an alternative y and a subset T of S containing x such that y is uniquely chosen.

Axiom 3.6.2. *(Dominating Alternative) If $x \notin C(S)$, then there exist $y \in C(S)$ and $T \subseteq S$ containing x such that $C(T) = y$.*

To see why Axiom 3.6.2 holds, suppose x is not chosen when the choice set is S. Then, x is not the best element in $N_x(S)$. Suppose the best element in $N_x(S)$ is y, and let $T = N_x(S) \subseteq S$. Then, since T is a connected set y must be uniquely chosen when the choice set is T.

Axiom 3.6.3 is similar to the standard expansion property. It says that if x is uniquely chosen when the choice set is T, y is uniquely chosen when the choice set is S, and T and S have a nonempty intersection, then either x or y must be uniquely chosen when the choice set is $T \cup S$.

Axiom 3.6.3. *(Expansion) Suppose $C(T) = x$ and $C(S) = y$. If $T \cap S \neq \emptyset$, then $C(T \cup S) = x$ or y.*

To see why it holds, suppose x is chosen when the choice set is T and y is chosen when the choice set is S. In our model, this can only happen if T and S are connected sets. If T and S have a nonempty intersection, then $T \cup S$ must also be a connected set. Therefore, a unique element must be chosen when the choice set is $T \cup S$. Given that x is the best alternative in T, and y is the best alternative in S, the only possible choices are x and y.

The last property follows from an observation that given a network we can divide any connected set into two connected sets with a nonempty intersection.

Axiom 3.6.4. *(Separability) Suppose $|S| \geq 3$. If $C(S) = x$, then there exist non-singleton $T_1, T_2 \subsetneq S$ with $T_1 \cap T_2 \neq \emptyset$ and $T_1 \cup T_2 = S$ such that $C(T_1) = x$ and $C(T_2) = y$ for some $y \in S$.*

Suppose x is uniquely chosen when the choice set is S. Then, S must be a connected set. Given the network structure we can separate S into two connected sets, say T_1 and T_2, with a nonempty intersection. If x is in T_1, then x must be uniquely chosen when the choice set is T_1, and the best element in T_2 must be uniquely chosen when choice set is T_2.

Theorem 3.6.1 shows that Axioms 3.6.1-3.6.4 are necessary and sufficient to characterize the induced network choice.

Theorem 3.6.1. *A choice correspondence C satisfies Contraction, Dominating Alternative, Expansion, and Separability if and only if it is an induced network choice.*

Proof of Theorem 3.6.1

Necessity is obvious from the previous discussion. We prove sufficiency. Let xPy if there exists $S \supseteq \{x, y\}$ such that $C(S) = x$

Claim 3.6.1. *P is acyclic.*

Proof. Suppose $x_1 P x_2 P \cdots P x_n P x_1$. Then, there exist $S_1, \ldots S_n$ with $S_i \supseteq \{x_i, x_{i+1}\}$ for $i < n$ and $S_n \supseteq \{x_1, x_n\}$ such that $C(S_i) = x_i$. Consider the set $T = S_1 \cup S_2 \cup \cdots \cup S_n$. Note that, by Axiom 3.6.1, we cannot have $x_i \in C(T)$ since $x_i \in S_{i-1} \subseteq T$ for $i > 1$ and $x_1 \in S_n \subseteq T$, but $C(S_i) = x_i$. Furthermore, we cannot have $y \in C(T)$ for any $y \notin \{x_1, \ldots, x_n\}$ since $y \in S_i$ for some i, but $y \notin C(S_i)$. Hence, we cannot assign any alternative to $C(T)$. Therefore, P is acyclic. □

Let $\succ$ be a completion of P. Define $\mathcal{N}$ as

$$(x, y) \in \mathcal{N} \text{ if and only if } C(\{x, y\}) \text{ is a singleton}$$

and let $N_x(S)$ be given by

$$N_x(S) = \{y \in S | \, \exists \{x_1, \dots, x_k\} \subseteq S \text{ with } x_1 = x, x_k = y, \text{ and } (x_i, x_{i+1}) \in \mathcal{N} \text{ for } i < k\}$$

Claim 3.6.2. *Let $\{x_1, \dots, x_k\}$ be such that $C(\{x_i, x_{i+1}\})$ is a singleton for $i < k$. Then, $C(\{x_1, \dots, x_k\})$ is a singleton.*

Proof. Suppose $C(\{x_1, x_2\})$ and $C(\{x_2, x_3\})$ are singletons. Since $\{x_1, x_2\} \cap \{x_2, x_3\} \neq \emptyset$, by Axiom 3.6.3, $C(\{x_1, x_2, x_3\})$ is a singleton. Now suppose $C(\{x_1, \dots, x_j\})$ and $C(\{x_j, x_{j+1}\})$ are singletons. Since $\{x_1, \dots, x_j\} \cap \{x_j, x_{j+1}\} \neq \emptyset$, $C(\{x_1, \dots, x_j, x_{j+1}\})$ is a singleton. Iterating this, we get that $C(\{x_1, \dots, x_k\})$ is a singleton. □

Claim 3.6.3. *If $C(S) = y$, then $y \in N_x(S)$ for all $x \in S$.*

Proof. Notice that the claim is trivial if $S = \{x, y\}$. Suppose the claim is true for S with $|S| < n$. Let S with $|S| = n$ be given and suppose $C(S) = y$. By Axiom 3.6.4, there exist non-singleton $T_1, T_2 \subsetneq S$ with $T_1 \cap T_2 \neq \emptyset$ and $T_1 \cup T_2 = S$ such that $C(T_1) = y$ and $C(T_2) = z$. Pick $t \in T_1 \cap T_2$. By induction hypothesis, $y \in N_t(T_1)$ and $z \in N_t(T_2)$. Since $T_1, T_2 \subsetneq S$, by A.1, $y, z \in N_t(S)$. Now pick $x \in S$. Either $x \in T_1$ or $x \in T_2$. If $x \in T_1$, then by induction hypothesis, $y \in N_x(T_1)$, and by A.1, $y \in N_x(S)$. If $x \in T_2$, then by induction hypothesis, $z \in N_x(T_2)$, and by A.1, $z \in N_x(S)$. But then, $z \in N_t(S)$ and $z \in N_x(S)$. By A.2, $N_x(S) = N_z(S) = N_t(S)$. Since $y \in N_t(S)$ we should have $y \in N_x(S)$. □

Claim 3.6.4. $C(S) = \{x \in S | x = \text{argmax}(\succ, N_x(S))\}$

Proof. Firsty, suppose $x \in C(S)$. We show that $x = \text{argmax}(\succ, N_x(S))$. Pick $z \in N_x(S)$. By definition, there exists $\{x_1, \dots, x_k\} \subseteq S$ with $x_1 = x$, $x_k = z$ such that $C(\{x_i, x_{i+1}\})$ is a singleton for $i < k$. By Claim 3.6.2, $C(\{x_1, \dots, x_k\})$ is a singleton. Since $x \in C(S)$ and $x \in \{x_1, \dots, x_k\} \subseteq S$, by Axiom 3.6.1, $C(\{x_1, \dots, x_k\}) = x$. Therefore, xPz, and hence $x \succ z$.

Now suppose $x = \text{argmax}(\succ, N_x(S))$. We show that $x \in C(S)$. Suppose $x \notin C(S)$. Then, by Axiom 3.6.2, there exist $y \in C(S)$ and $T \subseteq S$ containing x such that $C(T) = y$. By definition of P, we have yPx and hence $y \succ x$. By Claim 3.6.3, $y \in N_x(T)$, and by A.1, $y \in N_x(S)$. This contradicts $x = \text{argmax}(\succ, N_x(S))$. $\square$

BIBLIOGRAPHY

[1] Aaker, J. L. (1997): "Dimensions of brand personality," *Journal of Marketing Research*, 34(3), 347–356.

[2] Agranov, M. and P. Ortoleva (2017): "Stochastic choice and preferences for randomization," *Journal of Political Economy*, 125(1), 40–68.

[3] Aguiar, V. H. (2015): "Stochastic choice and attention capacities: Inferring preferences from psychological biases," *Working paper*.

[4] ——— (2017): "Random categorization and bounded rationality," *Economics Letters*, 159, 46–52.

[5] Anderson, J. R. (1983): "A spreading activation theory of memory," *Journal of Verbal Learning and Verbal Behavior*, 22(3), 261–295.

[6] Anderson, J. R. and G. H. Bower (2014): *Human Associative Memory*, Psychology press.

[7] Anscombe, F. J. and R. J. Aumann (1963): "A definition of subjective probability," *Annals of Mathematical Statistics*, 34(1), 199–205.

[8] Apesteguia, J. and M. A. Ballester (2009): "A theory of reference-dependent behavior," *Economic Theory*, 40(3), 427–455.

[9] Baker, W., J. Hutchinson, D. Moore, and P. Nedungadi (1986): "Brand familiarity and advertising: effects on the evoked set and brand preference," *Advances in Consumer Research*, 13, 637.

[10] Baliga, S., E. Hanany, and P. Klibanoff (2013): "Polarization and ambiguity," *American Economic Review*, 103(7), 3071–83.

[11] Bar-Hillel, M. (1980): "The base-rate fallacy in probability judgments," *Acta Psychologica*, 44(3), 211–233.

[12] Barberis, N., A. Shleifer, and R. Vishny (1998): "A model of investor sentiment," *Journal of Financial Economics*, 49(3), 307–343.

[13] Benjamin, D. J. (2019): "Errors in probabilistic reasoning and judgment biases," in *Handbook of Behavioral Economics-Foundations and Applications 2*, ed. by B. D. Bernheim, S. DellaVigna, and D. Laibson, Elsevier, 69–186.

[14] Bernheim, B. D. and A. Rangel (2009): “Beyond revealed preference: choice-theoretic foundations for behavioral welfare economics,” *Quarterly Journal of Economics*, 124(1), 51–104.

[15] Binmore, K., L. Stewart, and A. Voorhoeve (2012): “How much ambiguity aversion?” *Journal of Risk and Uncertainty*, 45(3), 215–238.

[16] Brady, R. L. and J. Rehbeck (2016): “Menu-dependent stochastic feasibility,” *Econometrica*, 84(3), 1203–1223.

[17] Bronnenberg, B. J., J. B. Kim, and C. F. Mela (2016): “Zooming in on choice: How do consumers search for cameras online?” *Marketing Science*, 35(5), 693–712.

[18] Bucklin, R. E. and C. Sismeiro (2003): “A model of web site browsing behavior estimated on clickstream data,” *Journal of Marketing Research*, 40(3), 249–267.

[19] Camerer, C. (1995): “Individual decision making,” in *Handbook of Experimental Economics*, ed. by J. H. Kagel and A. E. Roth, Princeton University Press, 587–703.

[20] Caplin, A. and M. Dean (2011): “Search, choice, and revealed preference,” *Theoretical Economics*, 6(1), 19–48.

[21] Cattaneo, M. D., X. Ma, Y. Masatlioglu, and E. Suleymanov (2019): “A random attention model,” *arXiv:1712.03448*.

[22] Charness, G., E. Karni, and D. Levin (2013): “Ambiguity attitudes and social interactions: An experimental investigation,” *Journal of Risk and Uncertainty*, 46(1), 1–25.

[23] Chen, Y. and S. Yao (2016): “Sequential search with refinement: Model and application with click-stream data,” *Management Science*.

[24] Cherepanov, V., T. Feddersen, and A. Sandroni (2013): “Rationalization,” *Theoretical Economics*, 8(3), 775–800.

[25] Collins, A. M. and E. F. Loftus (1975): “A spreading-activation theory of semantic processing.,” *Psychological Review*, 82(6), 407.

[26] Darley, J. M. and P. H. Gross (1983): “A hypothesis-confirming bias in labeling effects,” *Journal of Personality and Social Psychology*, 44(1), 20–33.

[27] Dean, M., Ö. Kıbrıs, and Y. Masatlioglu (2017): “Limited attention and status quo bias,” *Journal of Economic Theory*, 169, 93–127.

[28] Dominiak, A., P. Duersch, and J.-P. Lefort (2012): “A dynamic Ellsberg urn experiment,” *Games and Economic Behavior*, 75(2), 625–638.

[29] Edwards, W. (1968): “Conservatism in human information processing,” in *Formal Representation of Human Judgment*, ed. by B. Kleinmuntz, Wiley, New York, 17–52.

[30] Ellsberg, D. (1961): "Risk, ambiguity, and the Savage axioms," *Quarterly Journal of Economics*, 75(4), 643–669.

[31] Epstein, L. G. (1999): "A definition of uncertainty aversion," *Review of Economic Studies*, 66(3), 579–608.

[32] ——— (2006): "An axiomatic model of non-Bayesian updating," *Review of Economic Studies*, 73(2), 413–436.

[33] Epstein, L. G. and M. Le Breton (1993): "Dynamically consistent beliefs must be Bayesian," *Journal of Economic Theory*, 61(1), 1–22.

[34] Epstein, L. G., J. Noor, and A. Sandroni (2008): "Non-Bayesian updating: A theoretical framework," *Theoretical Economics*, 3(2), 193–229.

[35] Epstein, L. G. and M. Schneider (2003): "Recursive multiple-priors," *Journal of Economic Theory*, 113(1), 1–31.

[36] Epstein, L. G. and J. Zhang (2001): "Subjective probabilities on subjectively unambiguous events," *Econometrica*, 69(2), 265–306.

[37] Filippis, R. D., A. Guarino, P. Jehiel, and T. Kitagawa (2016): "Updating ambiguous beliefs in a social learning experiment," *Working Paper, Centre for Microdata Methods and Practice*.

[38] Fishburn, P. C. (1970): *Utility Theory for Decision Making*, Wiley, New York.

[39] Fryer, R. G., P. Harms, and M. O. Jackson (2018): "Updating beliefs when evidence is open to interpretation: Implications for bias and polarization," *Working Paper*.

[40] Garber, J. L. L. (1995): "The package appearance in choice," *Advances in Consumer Research*, 22, 653–660.

[41] Gennaioli, N. and A. Shleifer (2010): "What comes to mind," *Quarterly Journal of Economics*, 125(4), 1399–1433.

[42] Gentner, D. and A. L. Stevens (2014): *Mental models*, Psychology Press.

[43] Ghirardato, P. (2002): "Revisiting Savage in a conditional world," *Economic Theory*, 20(1), 83–92.

[44] Ghirardato, P., F. Maccheroni, and M. Marinacci (2004): "Differentiating ambiguity and ambiguity attitude," *Journal of Economic Theory*, 118(2), 133–173.

[45] Ghirardato, P. and M. Marinacci (2002): "Ambiguity made precise: A comparative foundation," *Journal of Economic Theory*, 102(2), 251–289.

[46] Gilboa, I. and D. Schmeidler (1989): "Maxmin expected utility with non-unique prior," *Journal of Mathematical Economics*, 18(2), 141–153.

[47] ——— (1993): “Updating ambiguous beliefs,” *Journal of Economic Theory*, 59(1), 33–49.

[48] Goeree, M. S. (2008): “Limited information and advertising in the US personal computer industry,” *Econometrica*, 76(5), 1017–1074.

[49] Gul, F. and W. Pesendorfer (2014): “Expected uncertain utility theory,” *Econometrica*, 82(1), 1–39.

[50] ——— (2018): “Evaluating ambiguous random variables and updating by proxy,” *Working Paper, Princeton University*.

[51] Hanany, E. and P. Klibanoff (2007): “Updating preferences with multiple priors,” *Theoretical Economics*, 2(3), 261–298.

[52] Hansen, L. and T. J. Sargent (2001): “Robust control and model uncertainty,” *American Economic Review*, 91(2), 60–66.

[53] Hauser, J. R. and B. Wernerfelt (1990): “An evaluation cost model of consideration sets,” *Journal of Consumer Research*, 16(4), 393–408.

[54] Hoffman, D. L. and T. P. Novak (1996): “Marketing in hypermedia computer-mediated environments: Conceptual foundations,” *Journal of Marketing*, 60(3), 50–68.

[55] Honka, E. and P. Chintagunta (2016): “Simultaneous or sequential? search strategies in the US auto insurance industry,” *Marketing Science*, 36(1), 21–42.

[56] Horan, S. (2013): “Random consideration and choice,” *Mimeo, CIRPÉE and Université du Québeca Montréal*.

[57] Hunt, R. R. and H. C. Ellis (1999): *Fundamentals of Cognitive Psychology*, New York, NY, US: McGraw-Hill.

[58] Johnson, E. J., W. W. Moe, P. S. Fader, S. Bellman, and G. L. Lohse (2004): “On the depth and dynamics of online search behavior,” *Management Science*, 50(3), 299–308.

[59] Kahana, M. J. (2002): “Associative symmetry and memory theory,” *Memory & Cognition*, 30(6), 823–840.

[60] Kahneman, D. and A. Tversky (1972): “Subjective probability: A judgment of representativeness,” *Cognitive Psychology*, 9, 430–454.

[61] ——— (1973): “On the psychology of prediction,” *Psychological Review*, 80(4), 237.

[62] ——— (1979): “Prospect theory: An analysis of decision under risk,” *Econometrica*, 47(2), 263–292.

[63] Keane, M. T., M. O’Brien, and B. Smyth (2008): “Are people biased in their use of search engines?” *Communications of the ACM*, 51(2), 49–52.

[64] Kim, J. B., P. Albuquerque, and B. J. Bronnenberg (2010): "Online demand under limited consumer search," *Marketing Science*, 29(6), 1001–1023.

[65] Koulayev, S. (2010): "Estimating demand in online search markets, with application to hotel bookings," *Boston University*.

[66] Kovach, M. (2016): "Thinking inside the box: Status quo bias and stochastic consideration," *Working paper*.

[67] Kreps, D. M. (1988): *Notes on the Theory of Choice*, Westview Press: Underground Classics in Economics.

[68] Lleras, J. S., Y. Masatlioglu, D. Nakajima, and E. Y. Ozbay (2017): "When more is less: Limited consideration," *Journal of Economic Theory*, 170, 70–85.

[69] Lord, C. G., L. Ross, and M. R. Lepper (1979): "Biased assimilation and attitude polarization: The effects of prior theories on subsequently considered evidence," *Journal of Personality and Social Psychology*, 37(11), 2098–2109.

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

[71] Mandel, N. and E. J. Johnson (2002): "When web pages influence choice: Effects of visual primes on experts and novices," *Journal of Consumer Research*, 29(2), 235–245.

[72] Manzini, P. and M. Mariotti (2007): "Sequentially rationalizable choice," *American Economic Review*, 97(5), 1824–1839.

[73] ——— (2012): "Categorize then choose: Boundedly rational choice and welfare," *Journal of the European Economic Association*, 10(5), 1141–1165.

[74] ——— (2014): "Stochastic choice and consideration sets," *Econometrica*, 82(3), 1153–1176.

[75] Masatlioglu, Y. and D. Nakajima (2013): "Choice by iterative search," *Theoretical Economics*, 8(3), 701–728.

[76] Masatlioglu, Y., D. Nakajima, and E. Y. Ozbay (2012): "Revealed attention," *American Economic Review*, 2183–2205.

[77] Masatlioglu, Y. and E. A. Ok (2005): "Rational choice with status quo bias," *Journal of Economic Theory*, 121(1), 1–29.

[78] ——— (2014): "A canonical model of choice with initial endowments," *Review of Economic Studies*, 81(2), 851–883.

[79] McCracken, J. C. and M. C. Macklin (1998): "The role of brand names and visual cues in enhancing memory for consumer packaged goods," *Marketing Letters*, 9(2), 209–226.

[80] Meyer, D. E. and R. W. Schvaneveldt (1976): “Meaning, memory structure, and mental processes,” *Science*, 192(4234), 27–33.

[81] Minardi, S. and A. Savochkin (2017): “Subjective contingencies and limited Bayesian updating,” *HEC Paris Research Paper No. ECO-SCD-2017-1203*.

[82] Moe, W. W. (2003): “Buying, searching, or browsing: Differentiating between online shoppers using in-store navigational clickstream,” *Journal of Consumer Psychology*, 13(1), 29–39.

[83] Mullainathan, S. (2002a): “A memory-based model of bounded rationality,” *Quarterly Journal of Economics*, 117(3), 735–774.

[84] ——— (2002b): “Thinking through categories,” *Unpublished Paper*.

[85] Mullainathan, S., J. Schwartzstein, and A. Shleifer (2008): “Coarse thinking and persuasion,” *Quarterly Journal of Economics*, 123(2), 577–619.

[86] Munro, A. and R. Sugden (2003): “On the theory of reference-dependent preferences,” *Journal of Economic Behavior & Organization*, 50(4), 407–428.

[87] Odean, T. (1998): “Volume, volatility, price, and profit when all traders are above average,” *Journal of Finance*, 53(6), 1887–1934.

[88] Oestreicher-Singer, G. and A. Sundararajan (2012): “The visible hand? Demand effects of recommendation networks in electronic markets,” *Management Science*, 58(11), 1963–1981.

[89] Ortoleva, P. (2012): “Modeling the change of paradigm: Non-Bayesian reactions to unexpected news,” *American Economic Review*, 102(6), 2410–36.

[90] Pires, C. P. (2002): “A rule for updating ambiguous beliefs,” *Theory and Decision*, 53(2), 137–152.

[91] Rabin, M. (1998): “Psychology and economics,” *Journal of Economic Literature*, 36(1), 11–46.

[92] Rabin, M. and J. L. Schrag (1999): “First impressions matter: A model of confirmatory bias,” *Quarterly Journal of Economics*, 114(1), 37–82.

[93] Reutskaja, E., R. Nagel, C. F. Camerer, and A. Rangel (2011): “Search dynamics in consumer choice under time pressure: An eye-tracking study,” *American Economic Review*, 101(2), 900–926.

[94] Sagi, J. S. (2006): “Anchored preference relations,” *Journal of Economic Theory*, 130(1), 283–295.

[95] Salant, Y. and A. Rubinstein (2008): “(A, f): Choice with frames,” *Review of Economic Studies*, 75(4), 1287–1296.

[96] Samuelson, W. and R. Zeckhauser (1988): "Status quo bias in decision making," *Journal of Risk and Uncertainty*, 1(1), 7–59.

[97] Sen, A. K. (1971): "Choice functions and revealed preference," *The Review of Economic Studies*, 38(3), 307–317.

[98] Shafer, G. (1976): *A Mathematical Theory of Evidence*, Princeton University Press.

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

[100] Stahl, D. O. (2014): "Heterogeneity of ambiguity preferences," *Review of Economics and Statistics*, 96(4), 609–617.

[101] Strzalecki, T. (2011): "Axiomatic foundations of multiplier preferences," *Econometrica*, 79(1), 47–73.

[102] Sugden, R. (2003): "Reference-dependent subjective expected utility," *Journal of Economic Theory*, 111(2), 172–191.

[103] Suppes, P. and R. D. Luce (1965): "Preference, utility, and subjective probability," *Handbook of Mathematical Psychology*, 3, 249–410.

[104] Tversky, A. (1969): "Intransitivity of preferences.," *Psychological Review*, 76(1), 31–48.

[105] ——— (1972): "Elimination by aspects: A theory of choice.," *Psychological Review*, 79(4), 281.

[106] ——— (2004): *Preference, Belief, and Similarity: Selected Writings*, MIT Press, 193–402.

[107] Tversky, A. and D. Kahneman (1991): "Loss aversion in riskless choice: A reference-dependent model," *Quarterly Journal of Economics*, 106(4), 1039–1061.

[108] Tyson, C. J. (2013): "Behavioral implications of shortlisting procedures," *Social Choice and Welfare*, 41(4), 941–963.

[109] Wright, P. and F. Barbour (1977): "Phased decision strategies: Sequels to an initial screening," in *Multiple Criteria Decision Making*, ed. by M. K. Starr and M. Zeleny, Amsterdam: North-Holland Publishing Co. 91–109.

[110] Zhang, J. (2002): "Subjective ambiguity, expected utility and Choquet expected utility," *Economic Theory*, 20(1), 159–181.

[111] Zhao, C. (2017): "Pseudo-Bayesian updating," *Working Paper, The University of Hong Kong*.

[112] ——— (2018): "Representativeness and similarity," *Working Paper, The University of Hong Kong*.