Grabbing The Forbidden Fruit: Restriction-Sensitive Choice*

Alexis GHERSENGORIN† Niels BOISSONNET‡

October 28, 2022 [Please click here for the latest version]

Abstract

Restricting individuals' access to some goods may steer their desire toward their substitutes, a phenomenon known as the *forbidden fruit effect*. We propose and study a model of *restriction-sensitive choice* (RSC) that rationalizes such behaviors and that is compatible with the prominent psychological explanations: reactance theory and commodity theory. We show how ingredients of our model can be identified from choice data, namely, from choice reversals caused by the removal of options. We give an axiomatic characterization of RSC. We also derive a preference ordering over opportunity sets for agents whose final choices follow our procedure. Three applications are then analyzed. We show that our model can accommodate the emergence of conspiracy theories and the backlash of integration policy targeted toward minorities. We finally study a principal's delegation problem to an agent whose choice is an RSC. We find that the effect on the agent's welfare is ambiguous.

KEYWORDS: choice theory, violations of WARP, reactance theory, freedom of choice.

*This paper previously circulated with the title "Reactance: a Freedom-Based Theory of Choice". We are grateful to Miguel Ballester, Roland Bénabou, Francis Bloch, Yves Breitmoser, Franz Dietrich, Georgios Gerasimou, Paola Manzini, Marco Mariotti, Pietro Ortoleva, Eduardo Perez-Richet, Evan Piermont and Jean-Marc Tallon for helpful discussions and suggestions, as well as seminar participants at LEMMA Paris 2, Paris 8, BiGSEM, PSE, Sciences-Po, the Cognitive Economics Conference 2021, BRIC 2022, FUR 2022 and ESEM 2022 in Milan. A. Ghersengorin thanks ANR-17-CE26-0003 for its financial support.

†Paris School of Economics, Université Paris 1 Panthéon-Sorbonne. Email: a.ghersen@gmail.com.

‡Bielefeld University. Email: niels.boissonnet@gmail.com.

"Prohibitions create the desire they were intended to cure."

Lawrence Durell

Restricting an individual's feasible opportunities may steer their desire toward the prohibited opportunities or their substitutes. This phenomenon is known as the *forbidden fruit effect*, a reference to the episode in Genesis when God tells Adam and Eve that they are free to help themselves to any food in the Garden of Eden except the fruit from the tree of the knowledge of good and evil, which they finally eat (Levesque, 2018). The forbidden fruit effect has received empirical support in various contexts, such as the choice of environmentally harmful products, media choices, reluctance to follow a nudge policy, smoking decisions, alcohol intake, eating behaviors, etc.[1] Although many of these decisions may have important economic consequences, this has rarely been explored in economics.

The forbidden fruit effect generates choice behaviors that are incompatible with the canonical model of preference (or utility) maximization. According to the latter, an agent holds a fixed ranking over alternatives and chooses the option ranked the highest among any set of opportunities they might face. Accommodating the forbidden fruit effect entails relaxing this standard requirement and allowing for menu-dependent choices. Specifically, studying reactions to restrictions amounts to investigating violations of the "Independence of Irrelevant Alternatives" (IIA) (Chernoff, 1954; Sen, 1971, property α) triggered by the *removal* of opportunities.[2] Let us illustrate this with a field experiment studied by Mazis, Settle and Leslie (1973). In 1972, Miami-Dade county decided to forbid phosphate use for laundry. Despite its strong environmental

[1] For the choice of environmentally harmful products, see Mazis, Settle and Leslie (1973), see also the "rolling coal" movement in the US (in reaction to regulations of cars gas emissions, some drivers modified their engine at significant costs in order to pollute more). See Arad and Rubinstein (2018) for the reaction to nudges. For smoking decisions, see Pechmann and Shih (1999). For alcohol consumption, see Hankin et al. (1993). For eating behaviors, see Jansen, Mulkens and Jansen (2007); Jansen et al. (2008). For media choices, see Bushman (2006); Sneegas and Plank (1998); Varava and Quick (2015); Gosselt, De Jong and Van Hoof (2012).

[2] Property α is a weakening of the *Weak Axiom of Revealed Preferences* (Samuelson, 1938), that is necessary and sufficient to explain a single-valued choice function by the maximization of a linear order (see Sen, 1971).

rationales, this decision raised significant protests as well as unexpected reactions. For the sake of "American freedom", some consumers, among whom some were not buying phosphate-based detergent prior to the law, started buying it in neighbouring counties, smuggling it at extra cost and stockpiling the (now) precious product for the 20 years to come.[3] Formally, denoting by x the phosphate detergent in a neighbouring county, y the same product in Miami and z a phosphate-free detergent in Miami, the following choice reversal happens: z is chosen from the set $\{x, y, z\}$ while x is chosen over z once y is removed, i.e., in the menu $\{x, z\}$.

In this paper, we study a class of choice procedures, named *restriction-sensitive choice* (RSC), that account for the forbidden fruit effect (section 1.2). RSC can be seen as a four-stage process. First, the decision maker (DM) categorizes the set of options into *types* (e.g., horizontal differentiation). Second, options within types are ranked according to a *utility function* u, which represents the DM's intrinsic satisfaction, or material welfare (e.g., vertical differentiation). Third, within each type, the DM determines a *threshold* utility level, below which the options are evaluated by a *reaction function* v (which differs from u). Fourth, the choice is made by choosing among the top available elements from each type (according to u), where the top element is evaluated through v or u depending on whether it is above or below the threshold. To illustrate the model, consider three options x, y, z that are horizontally and vertically differentiated; namely, x is of the same *type* of product as y, but at a higher price; z is of another type. The DM has an intrinsic preference for z over the options of the other type. This is captured through the *utility function* u: $u(z) > u(y) > u(x)$. Therefore, z is chosen from the set $\{x, y, z\}$. However, when the access to options of the first type is restricted to the bad one (i.e., x), then the DM gets a further motive for choosing an option of this type (i.e., choosing x over z), which generates a forbidden fruit effect. This is captured through the the threshold of the first type, which is between $u(y)$ and $u(x)$, and the *reaction function* v such that $v(x) > u(z)$.

[3] As Mazis, Settle and Leslie (1973) showed, this astonishing effect on behavior was consistent with consumers' beliefs reversal: Miami consumers were, on average, more prone to praise phosphate detergent for its efficiency than their Tampa county neighbors.

We interpret this as v combining welfare and the additional desire created by restrictions.

The two prominent explanations of the forbidden fruit effect in psychology have been *reactance theory* (Brehm, 1966) and *commodity theory* (Brock, 1968), both of them being consistent with RSCs (section 1.3).[4] Reactance relates people's reaction to restriction or prohibition to their attitudes toward freedom of behavior. When they feel that a specific freedom of behavior is threatened, they experience psychological reactance, a motivational state toward the *restoration* of this lost or threatened freedom. With this in mind, in our model, each type of option subjectively embodies a specific freedom.[5] The threshold marks the minimal welfare requirement such that when only options below it are available, the DM perceives this as a threat to that particular freedom. The reaction function v therefore captures the propensity of the DM to restore a threatened freedom. Commodity theory predicts that the more a commodity is perceived as unavailable or requiring much effort to be obtained, the more it will be valued. According to this interpretation, a type gathers similar commodities and when only options below the threshold are available this makes salient the restriction on this type of option, thereby increasing their attractiveness.

We investigate the identification of our model (section 2). We define a notion of revealed reaction to restriction in the following way: when we observe a choice reversal such as $z = c\{x, y, z\}$ but $x = c\{x, z\}$, we say that x *reacts to the absence of* y (section 2.1). Our interpretation is that the removal of y creates an additional desire to choose x. We show that ingredients of an RSC are essentially pinned down by this revealed reaction relation (section 2.2). In particular, x reacts to the absence of y implies that x and y are of the same type and

[4]See Rosenberg and Siegel (2018) for a review on psychological reactance theory; Lynn (1991) for a review on commodity theory.

[5]Importantly, types are not postulated *a priori* and objectively observed but subjectively perceived by the DM and thus revealed by the analysis. Psychologists emphasize that reactance reflects an attempt to restore the loss of concrete freedoms, that is, freedoms to choose diverse types of option. "Contrary to some interpretations (e.g. Dowd, 1975), the freedoms addressed by the theory are not "abstract considerations," but concrete behavioral realities. If a person knows that he or she can do X (or think X, or believe X, or feel X), then X is a specific, behavioral freedom for that person." (Brehm and Brehm, 2013, p.12)

x is below the threshold; therefore, the types and the thresholds can be identified in this way. Furthermore, building on a uniqueness result for the utility and the reaction functions, we give a definition of *welfare improvement* for an RSC. We illustrate by means of examples how our welfare criterion differs both from the conservative one of Bernheim and Rangel (2007, 2008, 2009) and the preference identified from choice with limited attention by Masatlioglu, Nakajima and Ozbay (2012), hence contributing to the literature on welfare analysis under nonstandard individual choice.[6]

Our identification results rely on the particular structure of RSCs. Therefore, this naturally leads to the question of the falsifiability of our model. We thus give an axiomatic characterization (section 3). To that purpose, we first suitably relax IIA by requiring a standard Expansion axiom. Then, building on our definition of revealed reaction to restriction and a companion one, we state four axioms that capture consistency conditions on choices responsive to restrictions. We show that these five axioms fully characterize RSC. Importantly, other frequently observed phenomena generate similar choice patterns as the ones resulting from the forbidden fruit effect. In particular, the analysis of the attraction effect by Ok, Ortoleva and Riella (2015) is also based on choice reversals. However, our axioms are incompatible with theirs as long as some choice reversals are observed. Furthermore, RSC is a specific case of *choice with limited attention* (Masatlioglu, Nakajima and Ozbay, 2012). Yet, the interpretation is different and therefore so are the welfare predictions. Finally, RSC is also a specific case of a large class of choice procedures, popularized by the seminal paper by Manzini and Mariotti (2007), that sequentially apply two rationals. In particular, RSC is a *transitive shortlist method*, according to which choices are made by sequentially applying a pair of transitive preferences (Horan, 2016). The reverse is however not true.

In section 4, we take the point of view of a DM who behaves according to an RSC and ask how they would evaluate the freedom offered by the different sets of opportunities they might face; thus contributing to the literature on

[6]See Manzini and Mariotti (2012); Chambers and Hayashi (2012); Rubinstein and Salant (2012); Apesteguia and Ballester (2015); Grüne-Yanoff (2022).

freedom of choice (see Baujard, 2007, for a survey of the literature). Building on the series of papers by Pattanaik and Xu (1990, 1998, 2000), we axiomatize a criterion to rank menus, which simply counts the number of types from which sufficiently good options (i.e., above the threshold) are feasible. We argue that this ordering integrates considerations about similarities between options (see Pattanaik and Xu, 2000; Nehring and Puppe, 2002) and the role of the preferences of the agent (see Pattanaik and Xu, 1998), two aspects that have been studied separately in the literature.

We finally study three applications of our choice model (section 5). Two social phenomena have often been related to reactance and documented by the psychology literature, but they are not readily explained using existing (economic) models of choice. First, reactance is introduced as a possible determinant of the formation of conspiracy theories. To accommodate this phenomenon, we study how reactance impacts the DM's belief when she has to choose a biased source of information. By removing an unchosen moderately biased source, the DM might reverse their choice and choose a more biased source in the opposite direction. This can represent why, if a DM feels that some information is not accessible or hidden, they might end-up holding extreme beliefs or adhere to conspiracy theories. Second, reactance provides an explanation of why repressive policies towards minorities may generate backlash, as suggested by empirical evidence. Additionally, it provides an argument for the evolutionary efficiency of reactance and its persistence in the long run. Finally, we introduce RSC in a principal-agent's setting. We study a typical delegation problem: a principal can constrain the decision set of an informed but biased agent, but cannot commit to contingent monetary transfers. In addition to the standard model (e.g. Alonso and Matouschek, 2008), the agent behaves according to an RSC. We find that this modifies the optimal delegation strategy. Either it forces the principal to restrict even more the set of allowed actions to prevent the agent from taking worse actions; or it forces the principal to allow the agent's preferred options. Hence the effect of reactance on the agent's material welfare is ambiguous. This depends on the principal's payoff and prior distribution over the states of the world.

1 The Model

1.1 Preliminaries

We work with a finite set of options X and denote by $\mathcal{X} = 2^X \setminus \emptyset$ the collection of non-empty subsets of X. Elements of $\mathcal{X}$ stand for the menus of options available to the DM. A **choice function** $c : \mathcal{X} \longrightarrow X$ associates to each menu the option chosen by the DM in this menu.[7] Namely, for any menu A, $c(A) \in A$.[8]

We are interested in studying the effect of restrictions of the set opportunities on the DM's choices. Hence, our interpretation is that the menu is exogenously given to the DM, who must then choose an option within this menu. We however do not explicitly model an agent who actually restrains the DM's set of opportunities, except in some applications.

Let us finally stress that options are defined by objective features that can incorporate contextual properties — for instance, in our introductory example, we differentiated the phosphate laundry in a supermarket in Miami from the same product in a supermarket in a neighbouring county. Yet, we need not formalize these objective features, we only require the observer to be able to distinguish the different options. As it will become clear, some of these features may matter for the DM's subjective categorization of options and thus will be revealed through the choices.

1.2 Restriction-Sensitive Choice

In this section, we state our model, then detail the choice procedure it induces and finally give the two main interpretations. We first introduce two defini-

[7] We focus on choice functions for the sake of simplicity: dealing with choice correspondences would add another layer of complexity that, we think, is not necessarily relevant in the present context. Nonetheless, we conjecture that, with an appropriate weakening of Sens' property alpha, our results would extend to choice correspondences.

[8] For simplicity, if we enumerate a set $\{x_1, \ldots, x_k\}$, we write $c\{x_1, \ldots, x_k\}$ instead of $c(\{x_1, \ldots, x_k\})$.

tions.

Definition 1. *The* **order induced** *by a function* $f : X \longrightarrow \mathbb{R}$ *is the complete and transitive binary relation* $\succsim^f \subseteq X^2$ *such that, for any* $x, y \in X$, $x \succsim^f y \iff f(x) \geq f(y)$.

Definition 2. *A function* $f : X \longrightarrow \mathbb{R}$ *is* **single-peaked** *with respect to the linear order* $\succ \subseteq X^2$*, if for any* $x, y, z \in X$ *such that* $x \succ y \succ z$, $v(y) \geq \min\{v(x), v(z)\}$.[9]

Equipped with these two definitions, we now state the definition of our model, *restriction-sensitive choice*.

Definition 3. *A choice function* c *is a* **restriction-sensitive choice** *(RSC) if there exist a partition* $\mathcal{T}$ *of the options into* types*, a* threshold $\lambda_T \in \mathbb{R}$ *for each* $T \in \mathcal{T}$*, a* utility function $u : X \longrightarrow \mathbb{R}$ *that induces a linear order on each* $T \in \mathcal{T}$ *and a* reaction function $v : X \longrightarrow \mathbb{R}$ *, such that:*

(i) for any menu A,

$$\{c(A)\} = \arg\max_{x \in d(A)} v(x), \tag{1}$$

where:

$$d(A) = \bigcup_{T \in \mathcal{T}} \arg\max_{x \in T \cap A} u(x);$$

(ii) for any $T \in \mathcal{T}$, $u(.) = v(.)$ *on* $\{x \in T \mid u(x) \geq \lambda_T\}$;

(iii) for any $T \in \mathcal{T}$, v *is single-peaked with respect to the order induced by* u *on* $\{x \in T \mid u(x) < \lambda_T\}$.

In this case, we say that $\langle \mathcal{T}, \{\lambda_T\}_T, u, v\rangle$ *is an* **RS-structure** *that* **rationalizes** *the choice function* c.

According to RSC, options are partitioned into types. Choices are made sequentially: first, the DM retains the best available options from each type

[9] A *linear order* is a complete, transitive and antisymmetric binary relation.

according to u, forming the set $d(A)$; then, the DM chooses among this set according to v. Points (ii) and (iii) specify how the two criteria u and v are related to each other. Namely, the options above the thresholds λ_T's are evaluated according to u in both stages of the maximization; while the options below are evaluated in the second stage by v, which can differ from u. We now give interpretations of the model. Comparisons to existing models are relegated to section 3.

1.3 Interpretations

We give two interpretations of RSC: one in terms of reactance, the other in terms of saliency of a prohibition and the additional attractiveness it generates.

Reactance: a freedom-based theory of choice. According to the psychology literature, reactance is a reaction of an individual to a threat to their freedom of behavior, that aims to *restore* the eliminated freedom. In our framework, potential threats to freedom are captured by restrictions of the opportunity set.

The DM categorizes the options into types, forming a partition. In the words of psychologists, a type represents a certain *behavioral freedom* (see Brehm and Brehm, 2013). Within each type, options are ranked and chosen according to an instrumental criterion, the utility function u. Our interpretation is that u represents the intrinsic satisfaction, or material welfare, of the agent. A clear example is when a type contains the same good but obtained or consumed through different channels. For instance, buying phosphate laundry in the supermarket next-door is less costly than getting it in a supermarket in a neighbouring county, but both options may be perceived as similar.

The thresholds represent the minimal welfare requirements for the DM's freedoms. Namely, as long as options with a satisfaction level at least as good as λ_T are available, no freedom concern is activated regarding the specific freedom embodied by T. This is captured by (ii) in the definition of an RSC, which implies that, in the second step maximization, the evaluation of those options are not distorted. On the contrary, if the best option available from T is below

λ_T, the DM deems that they do not have access to a sufficiently good option regarding the freedom associated to T. They may then be prone to react by being even more willing to choose an option from T, although this option gives a lower satisfaction. This is how they "restore" the eliminated freedom. It is captured through the function v used in the second maximization, which combines welfare and the propensity to react to a freedom limitation. Indeed, as it will become clear in section 2, to generate reactance, v must exceed u for options below λ_T.

Point (iii) imposes a specific shape of the reaction function v with respect to the utility function u. This reflects the DM's increasing willingness to react, up to a certain point, as the limitations on their freedom is tightened (see Rosenberg and Siegel, 2018, for evidence of this phenomenon). In RSC, it is captured through the fact that the less welfare the DM can obtain from a type, the more they are willing to react. Single-peakedness simply allows this to be true up to a certain point where welfare motives might weigh more in the trade-off between welfare and freedom, that is, there might be a point where the DM considers the welfare sacrifice to be too important.

Commodity theory: when salient prohibition increases desire. Commodity theory states that the value of objects for the individuals increase with their feeling that the objects are impossible of difficult to access (Brock, 1968).

According to this interpretation, a type gathers options that the DM considers as similar — e.g., providing similar consumption experience — but with different level of satisfaction, or material welfare (as captured by u). The threshold λ_T captures the level of satisfaction below which the restricted availability of options of type T becomes *salient*. In this case, the best alternative option available from T becomes a "forbidden fruit" (e.g., Levesque, 2018), and thus all the more attractive. This is captured by v which adds, on top of the welfare, an intrinsic pleasure of defying the restriction. Similarly to the interpretation in terms of reactance, point (iii) captures the idea that the additional desire may increase with the degree of the restriction, up to a certain point.

2 IDENTIFICATION AND WELFARE

2.1 Revealed Reaction to Restriction

We are interested in the effects of restrictions of the opportunity set on choice behavior. These effects are observed when the motivation created by the restriction — be it related to freedom or the intrinsic desire for forbidden objects — conflicts with other motives, such as welfare or material satisfaction.[10] Formally, they are revealed through choice reversals that are inconsistent with standard preference maximization; namely, through violations of the *Independence of Irrelevant Alternative* (IIA, or property α, Sen, 1971) triggered by the removal of options. In particular, we are interested in the DM's reaction to the deprivation of an "apparently irrelevant" option, hence the following definition.

Definition 4. *Let c be a choice function on $\mathcal{X}$ and $x, y \in X$. We say that x **reacts to the absence of** y, relative to c, if there exists z such that, $z = c\{x, y, z\}$, and $x = c\{x, z\}$. We denote it $x\mathbf{R}^c y$.*[11]

x reacts to the absence of y means that being deprived of the feasibility of y, the DM's motive to choose x is boosted. We show that the relation $\mathbf{R}^c$ allows to uniquely identify the ingredient of an RSC.

2.2 Identification of Restriction-Sensitive Choice

The first proposition characterizes the relation $\mathbf{R}^c$ for an RSC.

Proposition 1. *Suppose c is rationalized by the RS-structure $\langle \mathcal{T}, \{\lambda_T\}_T, u, v\rangle$. For any $x, y \in X$: $x\mathbf{R}^c y$, if and only if there exists $T \in \mathcal{T}$ and $z \notin T$, such that $x, y \in T$,*

[10]Following Brehm (1966), reactance is meaningful only when freedom conflicts with another motive. "Reactance is conceived to be a counterforce motivating the person to reassert or restore the threatened or eliminated freedom. It exists only in the context of other forces motivating the person to give up the freedom and comply with the threat or elimination." (Brehm and Brehm, 2013, p.37).

[11]Our results are the same if the relation $\mathbf{R}^c$ is not defined using a triplet, but any set, i.e.: $x\mathbf{R}^c y$ if there exists a set A such that $x = c(A \setminus \{y\}) \neq c(A) \neq y$.

$u(x) < u(y)$ and $v(x) > v(z) > v(y)$.[12]

A direct consequence of this proposition is the following corollary.

Corollary 1. *Suppose c is rationalized by the RS-structure $\langle \mathcal{T}, \{\lambda_T\}_T, u, v\rangle$. For any $x, y \in X$, if $x\mathbf{R}^c y$, then there exists $T \in \mathcal{T}$ such that $x \in T$ and $u(x) < \lambda_T$.*

These results corroborate our interpretations. Indeed, x reacts to the absence of y means that: x and y are considered by the DM as similar; y is intrinsically preferred to x ($u(x) < u(y)$); and being deprived of y boosts the DM's willingness to choose x ($v(x) > v(y)$). In terms of reactance, this restriction is perceived by the DM as a threat to their freedom. According to the second interpretation, removing x makes the restriction (more) salient and increases the desire to choose the alternative x.

Thanks to these results, we can define a specific kind of RS-structures that rationalize an RSC and whose elements are identified from this relation $\mathbf{R}^c$. For that purpose, let define for a choice function c, the set T_0^c of the options that are never involved in any reaction to restriction:

$$T_0^c = \{x \in X : \nexists\, y \text{ such that } x\mathbf{R}^c y \text{ or } y\mathbf{R}^c x\}.$$

Proposition 2. *Suppose c is an RSC. Then, there exists an RS-structure $\langle \mathcal{T}, \{\lambda_T\}, u, v\rangle$ that rationalizes c, such that:*

(i) $T_0^c \in \mathcal{T}$;

(ii) *for any* $T \in \mathcal{T}$, $\lambda_T = \min u\big(\{x \in T : \nexists\, y, x\mathbf{R}^c y\}\big)$.

In this case, we say that $\langle \mathcal{T}, \{\lambda_T\}, u, v\rangle$ is a **minimal RS-structure**.

Minimal RS-structures are appropriate to study RSCs for several reasons. First, it shows that options in T_0^c do not matter: they can be removed from every type and gathered together.[13] Second, the thresholds really capture the

[12] All proofs of this section can be found in Appendix A.

[13] The interpretation of T_0^c as a type must thus be qualified: these options do not particularly share common features, they simply do not fall in any category.

idea that as long as options giving a satisfaction of at least λ_T are available, the DM does not react to any restriction of the available options from T. Vice versa, when only options with satisfaction levels below the threshold are available, the DM's is prone to react to this restriction. Finally, any type besides T_0^c is *relevant*, in the sense that it gathers options that are related through the DM's responsiveness to restrictions. Namely, for any x in this type, it is related to some y in the same type through the relation $\mathbf{R}^c$. The following corollary makes these statements explicit.

Corollary 2. *Suppose c is rationalized by the minimal RS-structure $\langle \mathcal{T}, \{\lambda_T\}, u, v\rangle$ and define for each $T \in \mathcal{T}$, $x_T \equiv u^{-1}(\lambda_T)$. Then, for any $T \neq T_0^c$ and $x \in T$:*

(i) if $u(x) < \lambda_T$, then $x\mathbf{R}^c x_T$;

(ii) if $u(x) \geq \lambda_T$, then there exists $y \in T$ such that $y\mathbf{R}^c x$ and $y\mathbf{R}^c x_T$.

Another important aspect is that the types and the thresholds of minimal RS-structures are essentially unique.

Corollary 3. *Suppose c is rationalized by the minimal RS-structures $\langle \mathcal{T}, \{\lambda_T\}_T, u, v\rangle$ and $\langle \tilde{\mathcal{T}}, \{\tilde{\lambda}_T\}_T, \tilde{u}, \tilde{v}\rangle$. Then $\mathcal{T} = \tilde{\mathcal{T}}$ and $x_T = \tilde{x}_T$ for each T (where $x_T = u^{-1}(\lambda_T)$ and $\tilde{x}_T = u^{-1}(\tilde{\lambda}_T)$).*

Remark. In light of proposition 2, it is worth noting that the absence of choice reversals as in definition 4 is to be interpreted as a lack of traceable reaction to restriction. This does not necessarily mean that the DM has no concern for restrictions of their opportunity set. Rather, this means that these concerns (if any) are either too weak, or too aligned with their welfare to be identified as a force counterbalancing welfare.

We finally discuss the uniqueness of the functions u and v. Let c be rationalized by the RS-structure $\langle \mathcal{T}, \{\lambda_T\}_T, u, v\rangle$. What are the joint conditions on functions $\tilde{u}, \tilde{v}$ that ensures that $\langle \mathcal{T}, \{\lambda_T\}_T, \tilde{u}, \tilde{v}\rangle$ also represents c? One obvious sufficient condition is if there exists an increasing function $f : \mathbb{R} \longrightarrow \mathbb{R}$ such that $\tilde{u} = f \circ u$ and $\tilde{v} = f \circ v$. Now let suppose that there exist two functions $f, g : \mathbb{R} \longrightarrow \mathbb{R}$ such that $\tilde{u} = f \circ u$ and $\tilde{v} = g \circ v$. Denote the following set gathering all options that are above the threshold of their type:

(2) $$F \equiv \{x \in X : u(x) \geq \lambda_{T(x)}\}.$$[14]

One clear necessary condition is that f and g coincide and are increasing on $u(F)$, so that $f \circ u$ and $g \circ v$ satisfy points (i) and (ii) in definition 3. Because u is not directly used as a choice rule on options that are not in F, it is not necessary that f be increasing on $u(X)$. Yet, it represents choices within types, which implies that f is increasing on $u(T)$ for every T. Because v is ultimately the function through which choices are made, one might be tempted to say that g must be increasing on $v(X)$. This is however not exact because within types, the function v is never used to make choices. This problem does not arise as long as we impose one additional innocuous condition regarding the reaction function on certain pairs of options of similar types. This is captured by the following definition.

Definition 5. *An RS-structure $\langle \mathcal{T}, \{\lambda_T\}_T, u, v\rangle$ is an* **RS*-structure** *if, for any $T \in \mathcal{T}$ and any $x, y \in T$, such that $u(y) < u(x)$, $u(y) < \lambda_T$ and for all $z \notin T$, $v(y) > v(z) \implies v(x) > v(z)$, then $v(y) \geq v(x)$.*

The next proposition shows the existence of RS*-structure and that if we restrict ourselves to RS*-structures, the conditions regarding the utility and the reaction functions stated above are not only sufficient, but also necessary.

Proposition 3. *Let c be an RSC.*

(i) There exists an RS*-structure $\langle \mathcal{T}, \{\lambda_T\}_T, u, v\rangle$ that rationalizes it.

(ii) Furthermore, let $f, g : \mathbb{R} \longrightarrow \mathbb{R}$ be two real-mappings, $\langle \mathcal{T}, \{\lambda_T\}_T, f \circ u, g \circ v\rangle$ also rationalizes c if and only if f is increasing on $u(T)$ for every $T \in \mathcal{T}$, g is increasing on $v(X)$ and $f|_{u(F)} = g|_{u(F)}$.

2.3 Welfare

Our interpretation is that u captures the intrinsic preferences, or the welfare, of the DM. We however learn from proposition 3 that if c is rationalized by a minimal RS-structure with utility function u, having $u(x) > u(y)$ is not sufficient to conclude that x is welfare improving on y. It is sufficient if either x and y are in the same type, or both are above the thresholds (or a transitive closure of the latter ideas). Hence the following definition.[15]

Definition 6. *Suppose c is rationalized by a minimal RS-structure $\mathcal{S} = \langle \mathcal{T}, \{\lambda_T\}_T, u, v \rangle$. Then, for any $x, y \in X$, x is* **welfare improving** *on y, denoted $x \succ^w_{\mathcal{S}} y$, if either:*

(i) $T(x) = T(y)$ and $u(x) > u(y)$; or

(ii) $T(x) \neq T(y)$, $u(x) \geq \lambda_{T(x)}$ and there exists $z \in T(y)$, such that $u(z) \geq \lambda_{T(y)}$ and $u(x) > u(z)$.

The following proposition, which easily follows from corollary 3 and proposition 3, ensures that minimal RS-structures uniquely identify the welfare improving relation.

Proposition 4. *Suppose c is rationalized by the minimal RS-structures $\mathcal{S}$ and $\tilde{\mathcal{S}}$. Then $\succ^w_{\tilde{\mathcal{S}}} = \succ^w_{\mathcal{S}}$.*

Interestingly, this welfare criterion neither coincides with the model-free criterion of Bernheim and Rangel (2009) nor with the preference relation identified from a *choice with limited attention* (Masatlioglu, Nakajima and Ozbay, 2012) (see Appendix E for examples).

3 CHARACTERIZATION

In this section, we state our main theorem, which gives a full axiomatic characterization of RSCs, thereby showing the falsifiability of our model.

Restrictions are captured through the removal of options. Hence, expanding menus should reduce the motive to react. In particular, let x be chosen in menu A. Expanding A by adding a set of options B in which x is also chosen should not induce any change in the choice. Otherwise, the reversal from

[15]The focus on minimal RS-structure is justified by the fact that otherwise some options might be mis-categorized in a type while it is in T_0^c and some conclusions about welfare could be wrongly drawn.

$A \cup B$ to A would be triggered by the loss of an option from $B \setminus A$, which should prevent x from being chosen in B. This is what our first axiom, a standard relaxation of the *Weak Axiom of Revealed Preference*, imposes.[16]

Expansion (Exp). *For any $x \in X$, $A, B \in \mathcal{X}$, if $x = c(A) = c(B)$, then $x = c(A \cup B)$.*

Note that if x *reacts to the absence of* y, then **Exp** implies that $y = c\{x, y\}$. That is, for a reaction to restriction to be meaningful, it must trigger a choice of an even "worse" option than the one that is no more available. Furthermore, if z plays the same role as in definition 4, then **Exp** also implies that $z = c\{y, z\}$. Therefore, a typical pattern of reaction to restriction implies a binary choice cycle.

It is worth noting though that **Exp** prevents the following case. Let x, y, z be three options that are alphabetically ordered according to a dimension (e.g., political bias). The DM chooses z when the three options are available. But the DM wants to have access to the extreme option x so that when it is removed, y is chosen over z. However, when only x and y are available, the DM still prefers to choose y as it is closer to their first-best option. In some sense, we capture reactions to restriction that involve some minimal link to intrinsic satisfaction. In this case, if having access to x really matters so much to the DM, then they must chooose it in some cases, in particular above y.

Our next axioms involve the relation $\mathbf{R}^c$ and a companion one $\mathbf{P}^c$ that we now define. Assume we observe that *(i)* y is preferred to x, and *(ii)* for each pair z, t such that $t = c\{y, z, t\}$ and $y = c\{y, t\}$, we also observe that $t = c\{x, z, t\}$ and $x = c\{x, t\}$.[17] In this situation we suspect x to be as effective as y to react to any restriction that induces the DM to choose y. Yet, it need not be the case that $x\mathbf{R}^c y$, for there might be no third option z that allow to reveal a reversal

[16] It was already present in Sen (1971), named property γ, and was later used by Manzini and Mariotti (2007) under the name *Expansion*.

[17] Note that this is stronger than simply requiring that for any z such that $y\mathbf{R}^c z$ we also observe $x\mathbf{R}^c z$.

as stated by definition 4 — i.e. no z is chosen in $\{x, y, z\}$ while x is chosen in $\{x, z\}$. When this happens, we posit that x *potentially reacts* to the absence of y.

Definition 7. *Let c be a choice function on $\mathcal{X}$ and $x, y \in X$. We say that x **potentially reacts to the absence of** y, relative to c, if $y = c\{x, y\}$, there exists z, t such that $t = c\{y, z, t\}$ and $y = c\{y, t\}$, and for any such pair we also observe that $t = c\{x, z, t\}$ and $x = c\{x, t\}$. We denote it $x\mathbf{P}^c y$.*

We now state our axioms. Consider our introductory example. Assume that, when only an expensive phosphate-free detergent is available in their county, the DM reacts to the prohibition by going to the neighbouring county to get some phosphate detergent, while they stay in their own county when a cheap phosphate-free detergent is available. This reveals that "buying phosphate detergent in the neighbouring county" reacts to the absence of "buying phosphate detergent in Miami supermaket", though it is revealed only when the price of the phosphate-free detergent is high. Assume also that, while the DM prefers not to transgress the law when they can buy phosphate in the neighbouring county, they decide to go on the black market when the latter is forbidden, and that they do so whatever the price of the available phosphate-free detergent may be. This reveals that "buying phosphate detergent on the black market" reacts to the absence of "buying phosphate detergent in the neighbouring county". For these behaviours to be consistent, we would like to also observe that if phosphate is first banned in the neighbouring county and then in the DM's county, the DM goes on the black market, thus revealing that "buying phosphate detergent on the black market" also reacts to the absence of "buying phosphate detergent in Miami supermakets". Hence our first axiom requires $\mathbf{R}^c$ to be transitive — note that $\mathbf{P}^c$ is transitive by definition.

R-TRANSITIVITY (R-Tran). *For any $x, y, z \in X$, if $x\mathbf{R}^c y$ and $y\mathbf{R}^c z$, then $x\mathbf{R}^c z$.*

Let us stress that **R-Tran** imposes also transitivity in the *similarity* between options, that is, it prevents the following situation: x is sufficiently close to y and $x\mathbf{R}^c y$, y is sufficiently close to z and $y\mathbf{R}^c z$, but x and z are too different to consider the possibility of x reacting to the absence of z.

Because $\mathbf{R}^c$ and $\mathbf{P}^c$ are typically incomplete, we also require the negative transitivity of these relations. It requires that if xR^cy, then for any z that would be in-between x and y — i.e., $y = c\{y, z\}$ and $z = c\{z, x\}$ — , it must be that xR^cz or zR^cy.

R-NEGATIVE TRANSITIVITY (R-NTran). *For any $x, y, z \in X$, such that $y = c\{x, y\}$, $z = c\{y, z\} = c\{x, z\}$:*

(i) if $\neg[x\mathbf{R}^cy]$ and $\neg[y\mathbf{R}^cz]$, then $\neg[x\mathbf{R}^cz]$;

(ii) if $\neg[x\mathbf{P}^cy]$ and $\neg[y\mathbf{P}^cz]$, then $\neg[x\mathbf{P}^cz]$.

To motivate our next axiom, consider an option y that never reacts to the absence of any other option, but whose removal triggers reaction from the DM by choosing x — i.e. $x\mathbf{R}^cy$. This means that as long as the DM has access to y, they never react to some limitation of their opportunity set by choosing y. At the same time, removing y triggers some reaction and motivate them to choose x. Therefore, our interpretation is that y is never chosen because of restriction-related motives: either y satisfies the DM's freedom requirement (interpretation 1); or y as long as y the unavailability of options similar to y is not sufficiently salient to make it more attractive (interpretation 2). Consider a third option z that is chosen over y and such that also $x\mathbf{R}^cz$, then z should be even less chosen because of restriction-related motives. Our third axiom imposes two conditions in that direction. First, any option that reacts to the absence of z must also react to the absence of y. Conversely, any option that reacts to the absence of y might not be good enough to react to the absence of z, but at least z must be chosen over this option.

R-CONSISTENCY (R-Con). *For any $x, y, z \in X$ such that $x\mathbf{R}^cy$, $x\mathbf{R}^cz$, $z = c\{y, z\}$ and there exists no t such that $y\mathbf{R}^ct$, and for any $u \in X$:*

(i) if $u\mathbf{R}^cz$, then $u\mathbf{R}^cy$;

(ii) if $u\mathbf{R}^c y$, then $z = c\{u, z\}$.[18]

To motivate our last axiom, we extend the phosphate example. Suppose that both "buying phosphate on the black market" (x) and "buying phosphate in a neighbouring county" (z) react to the absence of "buying phosphate in Miami supermakets" (t). Add the third option "buying phosphate in a further county" (y): quite naturally, z is chosen over y, and assume further that both z and y are chosen over x. Suppose that the DM considers going on the black market as a reaction to the prohibition in a further county, that is, $x\mathbf{R}^c y$. Said differently, the DM's propensity to choose a phosphate detergent when x is the only one available is greater than when y is available. Because both x and z reacts to the absence of a common option t, then one would expect that similarly the DM's motive to choose a phosphate detergent when y is the only one available is greater than when z available. Hence our third axiom requires that y potentially reacts to the absence of z. The second point says that if in addition $x\mathbf{P}^c z$, that is, whenever the DM considers going in a neighbouring county as a reaction to a restriction, they would also consider going on the black market if necessary, the same conclusion, that is, $y\mathbf{P}^c z$, should follow even if we only observe $x\mathbf{P}^c y$ and not necessarily $x\mathbf{R}^c y$. This axiom imposes some sort of monotonicity in the way the DM reacts to restrictions, in the sense that it forbids any "gap" in their reaction. That is to say, if x, y, z are transitively ranked in binary choices and they are all sometimes chosen as a reaction to the deprivation of a common option t, then the magnitude of the motivation to react should evolve monotonically from z to x. Hence, as long as $xR^c y$ or xP^y, it must be that at least $yP^c z$.

R-MONOTONICITY (R-Mon). *For any $x, y, z \in X$, such that $z = c\{y, z\}$, $y = c\{x, y\}$:*

(i) if $x\mathbf{R}^c t$ and $z\mathbf{R}^c t$ for some $t \in X$, then [$x\mathbf{R}^c y \implies y\mathbf{P}^c z$];

[18]Point *(ii)* can alternatively be seen as requiring that $u\mathbf{R}^c y$ cannot be revealed through the choice with z, hence $z = c\{u, z\}$, which is consistent with the interpretation that y and z are similar and z offers a better satisfaction.

(ii) if $x\mathbf{P}^c z$, then [$x\mathbf{P}^c y \implies y\mathbf{P}^c z$].

We now can state our representation theorem, according to which an RSC is entirely characterized by these five axioms.

Theorem 1. *A choice function c is an RSC if and only if it satisfies **Exp**, **R-Tran**, **R-NTran**, **R-Con** and **R-Mon**.*[19]

Comparisons to existing models. There exist several models that explain similar choice patterns as the ones generated by reactions to restrictions, in particular the choice reversal exhibited in definition 4 (see Manzini and Mariotti, 2007, 2012; Cherepanov, Feddersen and Sandroni, 2013; Masatlioglu, Nakajima and Ozbay, 2012; Ehlers and Sprumont, 2008; Ok, Ortoleva and Riella, 2015; Lleras et al., 2017; Apesteguia and Ballester, 2013; Horan, 2016; Ridout, 2021, among others).

In particular, a phenomenon frequently observed and studied is the attraction effect. This is the main focus of Ok, Ortoleva and Riella (2015). In particular, their definition of *revealed reference* is based on similar choice patterns: $z = c\{x, y, z\}$ and $x = c\{x, z\}$. Their interpretation is however significantly different: they argue that z beats x only with the "help" of y. Hence, while they interpret these reversals as revealing a relationship between y and z, we interpret it as revealing a relationship between x and y.[20] Their model and ours are actually incompatible. Indeed, as we noted after the statement of **Exp**, observing $x\mathbf{R}^c y$ and satisfying **Exp** imply a binary choice cycle, which is prevented by their *No-Cycle Condition*.

Among the different models that generate similar choice reversals, two have attracted a lot of attention: the *Rational Shortlist Method* (RSM) by Manzini and Mariotti (2007) and the *Choice with Limited Attention* (CLA) by Masatlioglu, Nakajima and Ozbay (2012). Masatlioglu, Nakajima and Ozbay (2012) actually show that these two models are both descriptively and behaviorally distinct. It happens that RSC is a special case of both CLA and RSM. First,

[19] The proof is in Appendix B.

[20] More precisely, the application of their definition identifies y as a *revealed reference* of z.

our operator $d(.)$ satisfies the unique condition of an *attention filter*; namely, for any A $d(A) = d(A \setminus \{x\})$ whenever $x \notin d(A)$. Let c be an RSC, given that the choice $c(A)$ follows from the maximization of the function v over the set $d(A)$, this shows that c is a CLA. Note however that our interpretations and thus our welfare conclusions are different (see section 2.3).

Second, let c be an RS rationalized by $\rangle\mathcal{T}, \{\lambda_T\}_T, u, v\langle$ and define the two orders P_1 and P_2 in the following way: $xP_1y \iff T(x) = T(y) \wedge u(x) > u(y)$; $xP_2y \iff v(x) > v(y)$. In that case, for any menu A, $c(A) = \max(\max(A, P_1), P_2)$. That is, the DM chooses as if she first keeps only options that are the best in each available type, and second, she chooses the best remaining one according to the binary comparisons. Therefore, (P_1, P_2) *sequentially rationalize* c.[21]

4 MEASURING FREEDOM

In this section, we adopt the interpretation in terms of reactance. We take the point of view of a DM whose final choices are rationalized by an RS-structure and ask how would they evaluate the freedom offered by the different sets of opportunity they might face. Starting with Jones and Sugden (1982) and Pattanaik and Xu (1990), there has been an important literature about freedom of choice that has proposed a wide variety of freedom measures based on the ranking of opportunity sets (see (see Baujard, 2007, for a survey of this literature). Importantly for us, two dimensions have been pointed out as relevant to the agents' valuation of their freedom: their (potential) preferences over options (see Pattanaik and Xu (1998) — henceforth PX98 — and Sen (1993)) and the similarity between different options (see Pattanaik and Xu (2000) — henceforth PX00 — and Nehring and Puppe (2002)).

According to a (minimal) RS-structure, it is through the interaction between the types and the thresholds that freedom concerns impact choices. This suggests that these two channels should impact the DM's assessment of freedom

[21]More precisely, the two orders are transitive, hence (P_1, P_2) is a *transitive shortlist method* (Horan, 2016).

offered by a given menu. The types represent classes of similar options,[22] suggesting that adding options of a similar type should not increase the DM's freedom of choice. In addition, the thresholds represent the DM's freedom demands. Hence, it seems natural that adding options increases the DM's valuation of freedom only if it gives access to items that satisfy this requirement, that is, above the threshold.

We characterize with two axioms a rule that reflects these arguments. As before, we denote X a finite set of options and $\mathcal{X} := 2^X \setminus \emptyset$ the collection of menus of options in X. Let $\langle \mathcal{T}, F, u, v \rangle$ be a minimal RS-structure defined on X and define F as in (2). Finally, $\succsim$ is a complete and transitive binary relation defined on $\mathcal{X}$.

To state our two axioms, we need to introduce the following definition. A menu A is **richer than** a menu B if for any $T \in \mathcal{T}$, if $T \cap F \cap A = \emptyset$, then $T \cap F \cap B = \emptyset$. So A is richer than B means that any type from which no element in F is available in A must also have no feasible options in $F \cap B$. Furthermore we say that A is **strictly richer than** B if A is richer than B but the reverse is not true.

Our first axiom says that (strictly) richer sets are always (strictly) preferred and imposes that it is an equivalence for singletons.

R-DOMINANCE.

(i) For any $A, B \in \mathcal{X}$: if A (strictly) richer than B, then $A (\succ) \succsim B$;

(ii) For any $x, y \in X$: $\{x\} \succ \{y\} \implies \{x\}$ strictly richer than $\{y\}$.

Note that part (i) of the axiom implies monotonicity in the sense of Kreps (1979): for any $A, B \in \mathcal{X}$, $A \supseteq B \implies A \succsim B$. Indeed, in this case, A is trivially richer than B. Part (ii) is an adaptation of Pattanaik and Xu (1990)'s *Indifference Between no Choice Situations*, which simply imposes an indifference between every singleton. They argue that singletons do not offer any freedom

[22] It is actually a specific case of PX00's analysis where the equivalence classes induced by the similarity relation form a partition.

of choice, hence they cannot be strictly ranked. This is still true in our case, except if only one the two options is above the threshold of its type (i.e., in F), which is exactly what is implied by (ii).

Our second axiom is an adaptation of the composition axioms used in Pattanaik and Xu's series of papers.

R-COMPOSITION. *For any $A, B, C, D \in \mathcal{X}$, such that $A \cap C = B \cap D = \emptyset$, $C \subseteq T$ and $D \subseteq T'$ for some $T, T' \in \mathcal{T}$, and A is not richer than C: if $A \succsim B$ and $C \succsim D$, then $A \cup C \succsim B \cup D$.*

Combining menus that do not overlap should preserve the ranking. This is however true only if combining really provides additional freedom, which is captured by the requirement that A is not richer that C (see PX00 for a complete discussion of their axiom).

For any menu A, we define $\Phi(A) = \{\mathcal{T}(x) \cap F \cap A | x \in A\}$, the collection of subsets containing every option of one type that is above the threshold and available in A. We can now state our representation theorem (the proof is in Appendix C).

Theorem 2. *$\succsim$ satisfies* R-DOMINANCE *and* R-COMPOSITION *iff for any menu A and B:*

$$A \succsim B \iff \#\Phi(A) \geq \#\Phi(B). \tag{3}$$

The interpretation is the following: what matters for the DM is to have access to more options, but only dissimilar objects — as captured by the distinct types — are valued. On top of that, within a certain class of similar options, the DM demands a minimal level of satisfaction to meet her freedom requirements, which is captured by the set F.

This measure is close to PX00's one. In addition to their representation, there is a role for preferences in this evaluation that is captured through the set F. Although PX98 also incorporate preferences, let us stress the key difference. Their starting point is a collection of possible preferences (i.e. complete and transitive orderings over the options) that a reasonable person may have. The resulting measure simply counts in a menu the number of options that are a maximiser of at least one of these preferences over the given menu. This approach integrates preferences relatively to a menu, simply attributing values to options that *could be* chosen in this menu. In contrast, in our approach, preferences are integrated in a more absolute way, in the following sense: below a certain level of satisfaction, even though the DM will have to choose an option, he does not attribute any freedom value to these potentially chosen items.[23] Even more, keeping the RSC in mind, some options that might be chosen later on, simply because of reactance, will not matter in the assessment of freedom, while some unchosen ones will matter.

5 Applications

We explore the scope of applicability of RSCs. We first show how our model can give plausible explanations to two observed and empirically supported phenomena: the formation of extreme beliefs — what we will call conspiracy theories — and the backfire effect of integration policies targeted toward minorities, two phenomena that have been related to reactance. We finally study the problem of a principal who must delegate a decision to a better-informed but biased agent who chooses according to an RSC.

5.1 Conspiracy Theories

As Sensenig and Brehm (1968) suggest, reactance has its counterpart in the realm of beliefs; namely the boomerang effect for psychologists (Hovland, Janis and Kelley, 1953) or the backfire effect for political scientists (Nyhan and Reifler, 2010; Wood and Porter, 2019).[24] In the wake of Covid 19 pandemics,

[23]To illustrate this, our measure can be equal to 0 on some non-empty menus, which is impossible either in PX98 or in PX00.

[24]The boomerang effect is "a situation in which a persuasive message produces attitude change in the direction opposite to that intended". The backfire effect is a concept from political science that refers to a situation in which evidence contradicting the subjects' prior belief

scholars argued such effects to be closely related to the formation of conspiracy theories and extreme beliefs (Adiwena, Satyajati and Hapsari, 2020).[25] We propose to accommodate this mechanism by adapting Che and Mierendorff (2019)'s single period model of attention allocation with reactance.

A DM must choose from two actions, l or r, whose payoffs depend on an unknown state $i \in \{L, R\}$. His prior belief that the state is R is denoted p and we assume that $p \in (0, 1/2]$. Before choosing his action, the DM acquires information. To that purpose, he can allocate his attention across four sources of information (e.g. newspapers). Two of them are *L-biased* and the two others are *R-biased*.

The sources are represented by statistical experiments, or signals. The L-biased ones, denoted σ^{LL} and σ^{L}, can only reveal the state R. Symmetrically, the R-biased ones, denoted σ^{RR} and σ^{R}, can only reveal the state L. For $i = L, R$, σ^{ii} is an *extreme* source, whereas σ^{i} is a *moderate* one, i.e. the former is more biased than the latter. Formally, σ^{i} sends signal s^{i} with probability 1 in state i and with probability $1 - \lambda$ in state $-i$, and σ^{ii} sends signal s^{i} with probability 1 in state i and with probability $1 - \delta$ in state $-i$. We assume that $3/4 > \lambda > \delta = 1/2$. The experiments induced by the moderate sources σ^{L} and σ^{R} are described in table 1. The signals σ^{LL} and σ^{RR} are obtained by replacing λ with δ.

σ^{L}

State/signal	s^{L}	s^{R}
L	1	0
R	$1 - \lambda$	λ

σ^{R}

State/signal	s^{L}	s^{R}
L	λ	$1 - \lambda$
R	0	1

Table 1: Experiments induced by the moderate sources.

Initially the DM faces the complete menu $M = \{\sigma^{LL}, \sigma^{L}, \sigma^{R}, \sigma^{RR}\}$. In terms of our representation, the set of L-biased sources and the one of R-biased sources each represent a type of options. For $i = L, R$, σ^{i} is strictly more Blackwell informative than σ^{ii}, therefore the DM will never choose any of the extreme

may reinforce their belief in the opposite direction.

[25] The fact that mass media did not give any credit to conspiracy theories has been pointed out as playing a role in reinforcing such theories through reactance.

sources when his opportunity set is M, that is: $d(M) = \{\sigma^L, \sigma^R\}$. The DM's demands from freedom are satisfied when the moderate sources are available, that is,$\lambda_L \leq u(\sigma^L)$ and $\lambda_R \leq u(\sigma^R)$. When facing the menu M, the DM foresees that his payoff from choosing action $a \in \{l, r\}$ in state $i \in \{L, R\}$ is u_a^i where : $u_r^R = u_l^L = 1$, $u_l^R = u_r^L = -1$. Hence the DM will prefer action r if and only if his posterior belief is greater than $1/2$. One can show that the DM's optimal allocation of attention is to choose the "own-biased news source"; namely the signal biased toward one's prior: in our case σ^L given that $p \leq 1/2$. The rationale for this is the following. The prior indicates action l as the optimal one. Hence, a breakthrough signal s^R from σ^L is more valuable than a breakthrough signal s^L from σ^R. And the biased signal s^L from σ^L is more aligned with the DM's prior belief than s^R from σ^R. Hence, he is better off allocating his attention to σ^L (see Che and Mierendorff, 2019, pp. 2999-3000, for the complete argument).

In the next period, the moderate R-biased source σ^R is no more available, either because the government actually banned this newspaper or simply because the DM perceives that this source is no longer existing: only L-biased or extremely R-biased ones are present. The DM now faces the menu $N = \{\sigma^{LL}, \sigma^L, \sigma^{RR}\}$. He interprets this removal as revealing that the disutility from making a mistake in state L — i.e. choosing action r — is lower than expected: he now foresees it to be $v_r^L = 0$. σ^{RR} is no more removed from consideration by σ^R, hence $d(N) = \{\sigma^L, \sigma^{RR}\}$. His utility from choosing σ^L is unchanged while the one attached to σ^{RR} is $v(\sigma^{RR}) = p + (1-p)\delta$ (for p sufficiently close to $1/2$ such that after signal s^R from σ^{RR}, the DM chooses action r).

As a consequence, some DMs with prior beliefs sufficiently close to $1/2$, who would have chosen news source σ^L in menu M, will choose the extreme source σ^{RR} in menu N and their default option becomes r.

Proposition 5. *There exists $p^\star < 1/2$ such that if $p \in [p^\star, 1/2]$:*

(i) The DM prefers σ^{RR} to σ^L in menu N;

(ii) After a realisation of signal s^R from σ^{RR}, the DM chooses action r.[26]

[26] All proofs of this section can be found in Appendix D.

This is in strong opposition as what would be obtained without reactance. Indeed, if the DM does not modify his utility when the menu shrinks, by removing σ^R, some DMs with prior belief strictly higher than $1/2$ would now choose the source σ^L instead and action l after a signal s^L.

5.2 Integration Policy Backlash

Can forced assimilation policy foster the integration of immigrants communities? While Alesina and Reich (2015)'s theory of nation building assumes that repressing the cultural practices of minorities spurs homogeneity, Bisin and Verdier (2001) suggest that the success of such policy may be mitigated by an increasing effort of parents to influence their children's cultural trait. In this application we show that, with reactance, one can even predict this policy to yield a backlash effect: the repressed immigrants react to repression by becoming more prompt to self-isolation. This additionally provides a rational to the persistence of reactance as an evolutionary efficient behavior.

Such a backslash effect has been recently documented by several papers. Some evidence suggests that the "burkha ban" in France in 2004 has strengthened the religious identity of French-Muslims (Abdelgadir and Fouka, 2020). Fouka (2020) shows that, in states which prohibited German Schools in the aftermath of World War I, German-Americans "were less likely to volunteer in World War II and more likely to marry within their ethnic group and to choose decidedly German names for their offspring".

To show how this backlash operates, we complement Bisin and Verdier (2001)'s account of cultural transmission with a reactance mechanism: as the repression increases, parents' educational freedom decreases and, reacting to this repression, they may endeavour to influence their children even more.[27] There are two cultural traits $\{m, M\}$ — for minority and Majority. The proportion of the minority q is assumed to be positive but lower than $1/2$. Each generation is composed of parents who have only one child. Intergenerational transmission

[27]For simplicity, we adopt a continuous setting, while our own framework is discrete. The ideas would be exactly the same with a discrete setting.

results from two socialization mechanisms. First, by vertical socialization the parents may directly transmit their cultural trait i with probability d^i. If, with probability $1 - d^i$, vertical socialization fails, then horizontal transmission occurs and the child adopts the traits of a random individual in society. Hence, the probability that a child from the minority be socialized by her parent's trait is:

$$P(d^i) \equiv d^i + (1 - d^i)q. \tag{4}$$

As Bisin and Verdier (2001), we argue that parents endeavour to influence their child. They have a unit of time to allocate between their effort to fix d^i — which costs $(d^i)^\beta$ unit of time, with $\beta > 1$ — and a leisure activity $t^i \in [0, 1]$, whose cost and utility are t^i. In addition, the government can implement a repressive policy $g^i \geq 1$ that may increase the parents' cost of influencing their child: a pair (t^i, d^i) costs $t^i + (d^i)^\beta g^i$ units of time for the parents. We posit that parents get a utility of 0 when their child is socialized to the other trait, while they get a utility $V(g^i)$ when she is socialized to their own trait. Hence, their expected utility of their child's socialization is $P(d^i)V(g^i)$. This means that, given a repressive policy, parents choose options $(t^i, d^i) \in [0, 1]^2$ from the menu

$$K_{g^i} \equiv \{(t^i, d^i) : t^i + (d^i)^\beta g^i \leq 1\},$$

to maximize

$$t^i + P(d^i)V(g^i), \tag{5}$$

In what follows, we also assume that V has the following shape:

$$V(g) = \begin{cases} \hat{V} & \text{if } \hat{g} \geq g \\ \hat{V}\frac{g^\lambda}{\hat{g}} & \hat{g} < g \end{cases}$$

For some $\hat{g} > 1$ with $\lambda > 1$ and $\hat{V} > 1$. Hence, after a threshold $\hat{g}$, the more repressive is the policy g, the greater is $V(g)$. The interpretation is that parents react to the repressive policy when they feel that their freedom to educate their child is threatened. In other words, more repression may create incentives to dedicate more resources to transmit their traits to their children. Note that λ represents some kind of reactance rate since as it increases, parents' willingness to influence their child also increases.

From the first order condition, we obtain that the unique equilibrium educational effort — the program (5) being concave — must satisfy:

$$d^{i\star}(g^i, q) = \left(\frac{1-q}{\beta}\frac{V(g^i)}{g^i}\right)^{\frac{1}{\beta-1}} \tag{6}$$

Given the shape of V, $d^\star$ strictly decreases with g on $(1, \hat{g})$ and strictly increases with g on $(\hat{g}, +\infty)$. In other words, when the repressive policy exceeds $\hat{g}$, the more repression, the more parents invest in having their child socialized to their own trait. This suggests that reactance is at work in this model. In the following lemma, we establish the precise connection between this model and our reactance framework.

Lemma 1. *The function C defined on $\{K_g\}_{g\geq 1}$, such that for all g*

$$C(K_g) = \{(t, d) \in K_g : (t, d) \text{ solves } (5)\}.$$

is a well-defined choice function and there exists an RSC C' defined on all compact subsets of $[0, 1]^2$ such that $C(K_g) = C'(K_g)$ for all $g \geq 1$.[28]

Assuming the repressive policy to solely concern the minority (i.e. $g^M = 1$), what does reactance imply for the population dynamics in this model? Let time $\tau \in [0, +\infty)$ be continuous and q_τ be the share of the population with the

[28]For convenience, we construct an RS-structure on this infinite collection of compact sets. Obviously, analogous results could be obtained by making the set of possible policies g and the menus K_g finite.

minority cultural trait at time τ. Then, we have[29]

$$\dot{q} = q(1-q)\Big(d^{m\star}\big(g^m, q\big) - d^{M\star}\big(1, 1-q\big)\Big).$$

Given (6), d satisfies the *cultural substitution property*.[30] This implies that q converges to some $q^\star \in (0,1)$, which satisfies $d^{m\star}\big(g^m, q\big) = d^{M\star}\big(1, 1-q\big)$ (see Bisin and Verdier, 2001, Proposition 1). Hence,

$$q^\star(g^m) = \frac{V(g^m)/g^m}{V(1) + V(g^m)/g^m} \tag{7}$$

Given that $V(g)/g$ increases with g when $g \in (\hat{g}, +\infty)$ this means that repressive policy increases the size of the minority. This prediction contrasts with Alesina and Reich (2015)'s suggestions.

Noting that reactance is presumably a characteristic cultural trait (Jonas et al., 2009), this model also provides a rationale for why reactance can be evolutionary efficient. Minorities which are more prompt to exhibit reactance are more likely to survive to repressive attempts to hinder their cultural practices.

To make precise this comparative statics statement, consider two minorities: one with a high reactance rate λ^H and one with a low reactance rate $\lambda^L < \lambda^H$. Denoting by $q^\star_L(.)$ and $q^\star_H(.)$ the equilibrium population share for these two minorities, the following proposition establishes that $q^\star$ is always higher for the high-reactance minority.

Proposition 6. *For all $g > \hat{g}$, $q^\star_H(g) > q^\star_L(g)$.*

5.3 Optimal Delegation and Reactance

We consider a typical delegation problem: a principal can constrain the decision set of an informed but biased agent, but cannot commit to contingent monetary transfers (see Holmstrom, 1980; Alonso and Matouschek, 2008, for a

[29]See Bisin and Verdier (2001, equation (3), footnote 9) for discussions about this differential equation.

[30]In Bisin and Verdier (2001, Definition 1), this property states that d is continuous, decreasing with q, and $d = 0$ when $q = 1$.

detailed review of the literature). In any organization (administrations, companies, etc.), many rules govern what agents can or cannot do, with the purpose of reducing agency costs incurred by principals while benefiting as much as possible from better-informed agents. One can think for instance of a head of a company who delegates stock management to plant managers, a regulator who delegates pricing decisions to a monopolist with unknown costs, or a manager who delegates pricing decision to sales persons.

Formally, a *principal* (she) has the legal right to take an action among a finite set $A = \{a^{LL}, a^L, a^R, a^{RR}\}$. The payoffs delivered by each action depends on the realization of a binary state of the world $\theta \in \{L, R\}$. While the principal only knows the probability $p \in [0, 1]$ that the state is R, an *agent* (he) is privately informed of the realization θ. The principal cannot use contingent transfers and must decide the set of actions among which the agent will choose.

Preferences. The principal's payoff for action a in state θ is the real number $\pi_\theta(a)$. Her preferred action is a^θ in state θ and her second favorite action $a^{\theta\theta}$. Her payoffs are written in table 2. The agent behaves according to an RS-structure with state-dependent utility and reaction functions. In both states, the types are $T^L = \{a^{LL}, a^L\}$ and $T^R = \{a^R, a^{RR}\}$ and $\lambda_L = u(a^{LL}), \lambda_R = u(a^{RR})$. The utility functions u_L, u_R and the reaction functions v_L, v_R are such that the agent reacts to the absence of $a^{\theta\theta}$ by choosing a^θ. In both states, he is more prone to restore the absence of a^{RR}. The functions are specified in table 3.

Principal	R	$\pi_R(a^R) > \pi_R(a^{RR}) > \pi_R(a^L) > \pi_R(a^{LL})$
	L	$\pi_L(a^L) > \pi_L(a^{LL}) > \pi_L(a^R) > \pi_L(a^{RR})$

Table 2: Principal's Payoffs.

Optimal Delegation. Denote $\mathcal{A} = 2^A \setminus \emptyset$ the set of *menus* of action. For any $M \in \mathcal{A}$, $a_\theta(M)$ is the (unique) action chosen by the agent in state θ when facing menu M. For any prior belief $p \in [0, 1]$, the objective of the principal is to solve

Agent	R	$v_R(a^R) > v_R(a^L) > u_R(a^{RR}) > u_R(a^{LL}) > u_R(a^R) > u_R(a^L)$
	L	$v_L(a^R) > v_L(a^L) > u_L(a^{LL}) > u_L(a^{RR}) > u_L(a^L) > u_L(a^R)$

Table 3: Agent's Utility and reaction functions.

the following maximization program, whose value is denoted $V(p)$:

(8) $$V(p) \equiv \max_{M \in \mathcal{A}} (1-p)\pi_L\big(a_L(M)\big) + p\pi_R\big(a_R(M)\big).$$

A **delegation strategy** is a mapping from the set of beliefs to the set of menus: $\sigma : [0,1] \longrightarrow \mathcal{A}$. If for any p, $(1-p)\pi_L\big(a_L(\sigma(p))\big) + p\pi_R\big(a_R(\sigma(p))\big) = V(p)$, we say that the delegation strategy σ is **optimal**.

We are interested in the effect of RSC on optimal delegation strategies by the principal, and consequently on the agent's welfare. Without any reaction to restrictions, given that the agent's interest is sufficiently aligned with the principal's ($u_R(a^R) > u_R(a^L)$ and $u_L(a^L) > u_L(a^R)$), for any prior $p \in [0,1]$, the optimal delegation is to let the agent choose among the set of actions $\{a^L, a^R\}$.[31] This strategy cannot be optimal with the RSC because the agent would always choose a^R and therefore, for p sufficiently close to 0, offering a^L as the only possible action is better for the principal. For moderate p, it might be better to let the agent choose among the whole set of actions (or equivalently among his preferred actions $\{a^{LL}, a^{RR}\}$) given that in state $\theta = L, R$, $a^{\theta\theta}$ is the second best action for the principal. It happens that it depends on the magnitude of the principal's payoffs, as summarized in proposition 7. Define the following beliefs:

$$\bar{p} = \frac{\pi_L(a^{LL}) - \pi_L(a^R)}{\pi_L(a^{LL}) - \pi_L(a^R) + \pi_R(a^R) - \pi_r(a^{RR})},$$

$$\underline{p} = \frac{\pi_L(a^L) - \pi_L(a^{LL})}{\pi_L(a^L) - \pi_L(a^{LL}) + \pi_R(a^{RR}) - \pi_R(a^L)},$$

$$\hat{p} = \frac{\pi_L(a^L) - \pi_L(a^R)}{\pi_L(a^L) - \pi_L(a^R) + \pi_R(a^R) - \pi_R(a^L)}.$$

[31] Here we assume that the utility functions u_L and u_R would be the ones used if the agent was not responsive to restrictions. Of course, this is a slight abuse of what we can identify from choices given proposition 3.

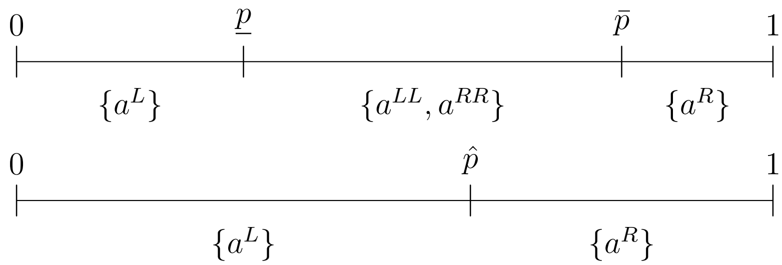

Figure 1: Optimal Delegation Strategies.

Proposition 7. *An optimal delegation strategy $\sigma^\star$ must induce the following actions.*

1. *If $\underline{p} < \bar{p}$:*

   (i) $a_L(\sigma^\star(p)) = a_R(\sigma^\star(p)) = a^L$ *for* $p < \underline{p}$;

   (ii) $a_L(\sigma^\star(p)) = a^{LL}$ *and* $a_R(\sigma^\star(p)) = a^{RR}$ *for* $\underline{p} < p < \bar{p}$;

   (iii) $a_L(\sigma^\star(p)) = a_R(\sigma^\star(p)) = a^R$ *for* $p > \bar{p}$;

   and it can induce either of the two possibilities respectively at the boundary beliefs $\underline{p}$ and $\bar{p}$.

2. *If $\underline{p} \geq \bar{p}$:*

   (i) $a_L(\sigma^\star(p)) = a_R(\sigma^\star(p)) = a^L$ *for* $p < \hat{p}$;

   (ii) $a_L(\sigma^\star(p)) = a_R(\sigma^\star(p)) = a^R$ *for* $p > \hat{p}$;

   and it can induce either of the two possibilities at the boundary belief $\hat{p}$.

Two possible optimal strategies are depicted in figure 1, implementing the actions described in proposition 7. In each strategy, the principal is indifferent between the two possible menus at boundary beliefs $\underline{p}, \bar{p}$ and $\hat{p}$. These strategies are the most direct ones, in the sense that each menu does not contain irrelevant actions that are never chosen by the agent. The logic behind this result is that RSC makes the agent's threat to choose bad actions (for himself) credible. Hence, the principal reacts either by constraining even more the agent's opportunity set; or on the contrary by offering him a greater satisfaction.

Agent's Welfare. If we measure the agent's material welfare through the utility functions u_L and u_R, we see from proposition 7 that the effect of the RSC is ambiguous. In the case where $\underline{p} \geq \bar{p}$, the effect is only negative, as the agent only has access to a unique action that is not among her best actions. But if $\underline{p} < \bar{p}$, then while there is still this negative effect when $p \leq \underline{p}$ or $p \geq \bar{p}$, on the contrary, for middle beliefs, reactance forces the principal to let the agent choose among her best options $\{a^{LL}, a^{RR}\}$.

Appendices

A Proofs of Section 2

Proof of proposition 1. Let c be an RSC rationalized by $\langle \mathcal{T}, \{\lambda_T\}, u, v\rangle$. The *if* part is left to the reader as it simple results from an application of the choice procedure of an RSC. We prove the *only if* part. Consider x, y, z such that $z = c\{x, y, z\}$ and $x = c\{x, z\}$, so $x\mathbf{R}^c y$. One can easily check that it is not possible that x, y, z are either all in the same type, or all in different types: in both cases, the choice results from the maximization of a unique function, so it must satisfy the *Weak Axiom of Revealed Preferences*. Hence exactly two among them must be of the same type, denote it T. Given that $z = c\{x, y, z\}$, this means that $z \in d\{x, y, z\}$ and z is the best element of its type according to u. Consequently, x is not of the same type as z as otherwise this would imply that $z = c\{x, z\}$. This also implies that $z \in d\{x, z\}$ and thus $v(x) > v(z)$. Hence $x \notin d\{x, y, z\}$, as otherwise it would imply $x = c\{x, y, z\}$, which is only possible if $x, y \in T$ and $u(y) > u(x)$. Therefore, $y \in d\{x, y, z\}$ and $z = c\{x, y, z\}$ implies that $v(z) > v(y)$. □

Proof of proposition 2. This simply follows from the proof of sufficiency of theorem 1. □

Proof of corollary 2. Point *(i)*. Let $\langle \mathcal{T}, \{\lambda_T\}_T, u, v\rangle$ a minimal RS-structure that rationalizes c, $T \neq T_0^c$ and $x \in T$. There must exist $x \in T$ such that $u(x) < \lambda_T$, as otherwise we would conclude from proposition 1 that $T = T_0^c$. Hence, there exists a sequence of options $(x_k)_{k=1}^n$ in T such that and $\lambda_T > u(x_1) > \cdots > u(x_n)$ and $\{x \in T : u(x) < \lambda_T\} = \{x_1, \ldots, x_n\}$.

From point (iii) in the definition of an RSC, there exists $k^\star \in \{1, \ldots, n\}$ such that $v(x_{k^\star}) = \max v\big(\{x_1, \ldots, x_n\}\big)$ and for any $k' < k < k^\star$, or $k' > k > k^\star$, $v(x_{k'}) < v(x_k) < v(x_{k^\star})$. Hence, it is sufficient to show that $x_1\mathbf{R}^c x_T$ and $x_n\mathbf{R}^c x_T$.

First, $x_1 \in T \neq T_0^c$ and $u(x_1) < \lambda_T$ mean that there exists $y \in T$ such that $x_1\mathbf{R}^c y$ as otherwise this would contradict the definition of λ_T from proposition 2. Hence, from proposition 1, there exists also $z \notin T$, such that $u(x_1) < u(y)$ and

$v(x_1) > v(z) > v(y)$. Furthermore, given the definition of x_1, $u(y) \geq \lambda_T$, hence $v(y) = u(y)$. Therefore, $v(x_1) > v(z) > \lambda_T = v(x_T)$ and $u(x_1) < \lambda_T = u(x_T)$, that is, $x_1 \mathbf{R}^c x_T$.

Second, given that $u(x_n) = \min u(T)$ and $x_n \in T \neq T_0^c$, there exists $y \in T$ such that $x_1 \mathbf{R}^c y$ as otherwise this would contradict the definition of T_0^c. Hence, there exists also $z \notin T$, such that $u(x_1) < u(y)$ and $v(x_1) > v(z) > v(y)$. This implies that $u(y) > u(x_{k^\star})$, which in turn implies that $v(y) \geq v(x_T) = \lambda_T$. We can conlude similarly as in the previous paragraph that $x_n \mathbf{R}^c x_T$.

Point *(ii)* directly follows from point *(i)* and propositions 1 and 2. □

Proof of proposition 3. The proof of *(i)* directly follows from the proof of sufficiency of theorem 1. The proof of the *if* part of *(ii)* is not complicated and thus left to the readers. We only prove the *only if* part.

The fact that f must be increasing on $u(T)$ for every $T \in \mathcal{T}$ simply follows from the fact that the function u represents the binary choices within each type. $f|_{u(F)} = g|_{u(F)}$ is a direct consequence of the requirement that utility and the reaction functions be equal on F.

We now prove that g must be increasing on $v(X)$. Suppose by contradiction that there exists $x, y \in X$ such that $v(x) > v(y)$ but $g \circ v(x) \leq g \circ v(y)$. Note that there must exist a type, denote it T, such that $x, y \in T$, as otherwise $v(x) > v(y) \implies x = c\{x, y\}$, which cannot be accommodated by $\langle \mathcal{T}, F, f \circ u, g \circ v \rangle$ if $g \circ v(x) \leq g \circ v(y)$. Define $x^\star = \arg\max v(T \setminus F)$ and x_T as in corollary 2.

(1) Consider the case where $u(x) > u(y)$. If $u(y) \geq \lambda_T$, this would mean that $x, y \in F$, in which case, given that $u|_F = v|_F$, $f|_{u(F)} = g|_{u(F)}$ and f is increasing on $u(T)$, it is impossible that $g \circ v(x) \leq g \circ v(y)$. If $u(y) \leq \lambda_T$, it means that $y \notin F$. The fact that $v(x) > v(y)$ implies that there exists $z \notin T$ such that $x = c\{x, z\}$ while $z = c\{y, z\}$, that is $v(x) > v(z) > v(y)$. Hence, $g \circ v(x) \leq g \circ v(y)$ cannot accommodate these choices.

(2) Consider the case where $u(x) < u(y)$. Then necessarily $u(x) < \lambda_T$, that is $x \notin F$. Given that $v(x) > v(y)$, there exists no $z \notin T$ such that $z = c\{x, z\}$ while $y = c\{y, z\}$. Conversely, if there exists $z \notin T$ such that $x = c\{x, z\}$ while $z = c\{y, z\}$, that is $v(y) > v(z) > v(x)$, then again $g \circ v(x) \leq g \circ v(y)$ cannot

accommodate these choices. If it is note the case, that is for every $z \notin T$ such that $x = c\{x, z\} \iff y = c\{y, z\}$, then $g \circ v(x) \leq g \circ v(y)$ does not satisfy the requirement of RS*-structure. $\square$

B Proof of Theorem 1

Proof of the necessity. Let $\mathcal{S} = \langle \mathcal{T}, \{\lambda_T\}_T, u, v\rangle$ be an RS-structure that rationalizes c. We denote $T(x)$ the type of the option x. Furthermore, for any $x, y, z \in X$, if $z = c\{x, y, z\}$ and $x = c\{x, z\}$, we write $x\mathbf{R}^c_z y$. Hence the definition of P^c can now be written: $xP^c y \iff$ there exists z, t such that $yR^c_t z$ and for any such pair $xR^c_t z$.

The following lemma shows that WARP is satisfied for each collection of menus that contains options of the same type.

Lemma 2. *For any $T \in \mathcal{T}$ and $A \subset B \subseteq T$, if $c(B) \in A$, then $c(A) = c(B)$.*

Proof. Let $c(B) = x$. Given that $B \subseteq T$, $x = c(B)$ implies that $c(B) = d(B)$, that is, $u(x) > u(y)$ for any $y \in B$, $y \neq x$. Because $A \subset B$, it means that $x = d(A)$, and therefore $x = c(A)$. $\square$

Exp. Let $x \in X$ and $A, B \in \mathcal{X}$ such that $x = c(A) = c(B)$. This means that $x \in d(A) \cap d(B)$. Hence, $u(x) > u(y)$ for all $y \in (A \cup B) \cap T(x)$, $y \neq x$, which implies that $x \in d(A \cup B)$. Moreover, $x = c(A) = c(B)$ implies that $v(x) > v(z)$ for all $z \neq x$ such that $z \in d(A) \cup d(B)$. Besides, $d(A \cup B) \subseteq d(A) \cup d(B)$, hence $v(x) > v(z)$ for all $z \neq x$, $z \in d(A \cup B)$. Hence, $x = c(A \cup B)$.

R-Tran. Let $x, y, z \in X$ such that $x\mathbf{R}^c y$ and $y\mathbf{R}^c z$. By definition,

$$x\mathbf{R}^c y \implies \exists t \in X, t = c\{x, y, t\} \text{ and } x = c\{x, t\},$$
$$y\mathbf{R}^c z \implies \exists t' \in X, t' = c\{y, z, t'\} \text{ and } y = c\{y, t'\}.$$

Proposition 1 implies that $T(x) = T(y) = T(z)$. Coupled with lemma 2, this shows that $T(t') \neq T(x) \neq T(t)$. Given that **Exp** is satisfied, we also know that $z = c\{z, y\}$ and $y = c\{y, x\}$. Hence, $u(z) > u(y) > u(x)$ and $v(x) > v(t) > v(y) > v(t') > v(z)$. Therefore, $t = c\{x, z, t\}$ and $x = c\{x, t\}$, which means that $x\mathbf{R}^c z$.

R-NTran (i). Let $x, y, z \in X$ such that $y = c\{x, y\}$, $z = c\{y, z\} = c\{x, z\}$, $\neg[x\mathbf{R}^c y]$ and $\neg[y\mathbf{R}^c z]$. Assume by contradiction that $x\mathbf{R}^c z$. Then there exists t such that $t = c\{x, z, t\}$ and $x = c\{x, t\}$ and, by proposition 1 and lemma 2, $T(x) = T(z) \neq T(t)$. Moreover, we have that $v(x) > v(t) > v(z)$ so that $v(x) > v(z)$. Hence, if $T(y) \neq T(x)$, then $z = c\{y, z\} = c\{x, z\}$ implies that

$$v(y) \underbrace{>}_{y=c\{x,y\}} v(x) > v(z) \underbrace{>}_{z=c\{y,z\}} v(y), \tag{9}$$

a contradiction. Now if $T(y) = T(x)$, then $u(z) > u(y) > u(x)$. But then either $v(y) > v(t)$, and then $y\mathbf{R}^c z$, or $v(t) > v(y)$ and then $x\mathbf{R}^c y$. In both cases we have a contradiction.

R-NTran (ii). Let $x, y, z \in X$ such that $y = c\{x, y\}$, $z = c\{y, z\} = c\{x, z\}$, $\neg[x\mathbf{P}^c y]$ and $\neg[y\mathbf{P}^c z]$. Assume by contradiction that $x\mathbf{P}^c z$. Hence, there exists t, u such that $z\mathbf{R}^c_u t$ and for any such t, u, $x\mathbf{R}^c_u t$. By proposition 1, $T(z) = T(t) = T(x) \equiv T$, $u(t) > u(z) > u(x)$ and $v(z), v(x) > v(t)$, which implies that $u(z) < \lambda_T$. Suppose that $y \notin T$, hence $v(z) > v(y) > v(x)$. Let t be such that both $xR^c t$ and $zR^c t$. Then it must be that $zR^c_y t$ but $\neg[xR^c_y t]$, which contradicts the fact that $xP^c z$. Suppose then that $y \in T$. This means that $u(z) > u(y) > u(x)$. Because $u(z) < \lambda_T$, then $u(y) < \lambda_T$. Given point (iii) in definition 3, we have that $v(y) > \min\{v(x), v(z)\}$. If $v(y) > v(z)$, then proposition 1 implies that $y\mathbf{P}^c z$, a contradiction. Hence $v(z) > v(y) > v(x)$. Let t, u be such that, $z\mathbf{R}^c_u t$, then $x\mathbf{R}^c_u t$, which means that $v(x) > v(u) > v(t)$, and therefore $v(y) > v(u) > v(t)$, and hence $y\mathbf{R}^c_u t$. This proves that $y\mathbf{P}^c z$, again a contradiction.

R-Con. Let $x, y, z \in X$ such that $x\mathbf{R}^c y$, $x\mathbf{R}^c z$, $z = c\{y, z\}$, and there exists no t such that $y\mathbf{R}^c t$. Proposition 1 implies that $T(z) = T(x) = T(y) \equiv T$ and $u(z) >$

$u(y)$. Furthermore, by proposition 2, it is without loss of generality to assume that $\mathcal{S}$ is minimal, and by proposition 1, $y\mathbf{R}^c t$ for no t implies that $u(y) \geq \lambda_T$, and hence $u(z) \geq \lambda_T$. Let $u \in X$ such that $u\mathbf{R}^c z$, so $u \in T$ and there exists $t \notin T$ such that $v(u) > v(t) > v(z) = u(z) > u(u)$. Hence $u(u) < \lambda_T \leq u(y)$. This means that $v(u) > v(t) > v(y) = u(y) > u(u)$. Hence, $u\mathbf{R}^c y$. This completes the proof of *(i)* in **R-Con**. Now assume that $u\mathbf{R}^c y$. This means that $u \in T$ and $u(z) > u(y) > u(u)$, hence $z = c\{x, z\}$, which proves *(ii)* in **R-Con**.

R-Mon. Let $x, y, z \in X$ such that $z = c\{y, z\}$, $y = c\{x, y\}$, $x\mathbf{R}^c t$, and $z\mathbf{R}^c t$ for some t. Assume that $x\mathbf{R}^c y$. By proposition 1, this means that $T(x) = T(y) = T(t) = T(z) \equiv T$, $\lambda_T > u(z) > u(y) > u(x)$, $v(y) < v(x)$. Point (iii) of definition 3 implies that $v(y) > v(z)$, from which we can conclude that $y\mathbf{P}^c z$. This proves *(i)*. Assume now that $x\mathbf{P}^c z$ and $x\mathbf{P}^c y$. By proposition 1, $T(x) = T(y) = T(z) \equiv T$ adn $\lambda_T > u(z) > u(y) > u(x)$. Point (iii) of definition 3 implies that $v(y) > \min\{v(x), v(z)\}$. If $v(y) > v(x)$, then for any t, u such that $z\mathbf{R}^c_u t$, given that $x\mathbf{R}^c_u t$, it must be that $y\mathbf{R}^c_u t$ and hence $y\mathbf{P}^c z$. Similarly, if $v(y) > v(z)$, $y\mathbf{P}^c z$ follows directly, which ends the proof of *(ii)*. □

Proof of the sufficiency. Let define the binary relation $\succ \subset X^2$ by $x \succ y$ if and only if $x = c\{x, y\}$ or $x = y$. It is clear that $\succ$ is complete and antisymmetric. For any transitive and complete binary relation $>$ defined on a set A, we write $\max(A, >) \equiv \{x \in A \mid x > y, \forall\, y \in A\}$. When $>$ is a linear order, with a slight abuse of notation, when no confusion can be made, we indifferently write $\max(A, >)$ for the singleton or for the element of the singleton. We define analogously $\min(A, >)$.

Lemma 3. *Let K a subset of X such that,*

$$\big((x, y) \in K^2 \iff \neg[x\mathbf{R}^c y] \text{ and } \neg[y\mathbf{R}^c x]\big), \tag{10}$$

then $\succ$ restricted to K^2 is a linear order and for all $K' \subseteq K$, $c(K) \succ y$ for all $y \in K'$.

Proof. Let K satisfying (10), $x, y, z \in K$ and $x \succ y \succ z$. Suppose by contradiction that $z \succ x$. If $x = c\{x, y, z\}$, then $z\mathbf{R}^c y$, which contradicts that $(y, z) \in K^2$

and K satisfies (10). The same reasoning applies if either $y = c\{x, y, z\}$ or $z = c\{x, y, z\}$. Hence, we conclude that $x \succ z$.

Moreover, let $K' \subseteq K$, the transitivity of $\succ$ on K, implies that there exists $x \in K'$ such that $x \succ y$ for any $y \in K'$. By **Exp**, $x = c(K')$. □

Define now

$$X^{\downarrow} = \bigcup_{y \in X} \{x \in X : x\mathbf{R}^c y\}$$
$$X^{\uparrow} = \bigcup_{y \in X} \bigcap_{t \in X} \{x \in X : y\mathbf{R}^c x, \neg[x\mathbf{R}^c t]\}$$

Let $\tilde{X} = X^{\uparrow} \cup X^{\downarrow}$ and for all $x \in X^{\downarrow}$, $R^{\uparrow}(x) = \{y \in X^{\uparrow} : x\mathbf{R}^c y\}$.

Lemma 4. *If $x \in X^{\downarrow}$, then $R^{\uparrow}(x) \neq \emptyset$.*

Proof. Let $x \in X^{\downarrow}$, i.e. $x\mathbf{R}^c y$ for some $y \in X$. If $y \in X^{\uparrow}$, this terminates the proof. Suppose that $y \notin X^{\uparrow}$, then there exists z_1 such that $y\mathbf{R}^c z_1$, which by **R-Tran** implies that $x\mathbf{R}^c z_1$. Either $z_1 \in X^{\uparrow}$, which ends the proofs, or there exists z_2 such that $z_1\mathbf{R}^c z_2$, which again by **R-Tran** implies that $x\mathbf{R}^c z_2$. At each step k, we replicate the same reasoning. Because X is finite, there must exist n such that for all $t \in X$, $\neg[z_n\mathbf{R}^c t]$, i.e., $z_n \in X^{\uparrow}$. Yet, **R-Tran** also implies that $x\mathbf{R}^c z_n$. Hence, $R^{\uparrow}(x) \neq \emptyset$. □

Note that lemma 3 implies that $\succ$ is transitive on $X^{\uparrow}$. Hence, lemma 4 implies the existence, for all $x \in X^{\downarrow}$, of $m(x)$, defined by:

$$m(x) \equiv \min(R^{\uparrow}(x), \succ). \tag{11}$$

Lemma 5. *For all $x, y \in X^{\downarrow}$, if $R^{\uparrow}(x) \cap R^{\uparrow}(y) \neq \emptyset$, then $m(x) = m(y)$;*

Proof. Let $x, y \in X^{\downarrow}$. Assume there exists $t \in R^{\uparrow}(x) \cap R^{\uparrow}(y)$ and let $t' = m(x)$. We show that $t' \in R^{\uparrow}(y)$. If $t = t'$ there is nothing to prove. If $t \neq t'$, then by definition of t' and since $t \in R^{\uparrow}(x)$, we have $t \succ t'$. Given that $t' \in X^{\uparrow}$ we have that $\neg[t'\mathbf{R}^c z]$ for any $z \in X$. Since $x\mathbf{R}^c t$, $x\mathbf{R}^c t'$ and $t \succ t'$, by **R-Con**(i) $y\mathbf{R}^c t$, implies that $y\mathbf{R}^c t'$, i.e. $t' \in R^{\uparrow}(y)$.

We prove symmetrically that $m(y)$ belongs to $R^{\uparrow}(x)$. Hence, by definition $m(x) \succ m(y)$ and $m(y) \succ m(x)$. Given that $\succ$ is a linear order on $x^{\uparrow}$, $m(x) = m(y)$. □

Since X is finite there exists n^* such that we can index the set $\{m(x) : x \in X^{\downarrow}\}$ by a sequence $(m(i))_{1 \leq i \leq n^*}$ such that $i \neq j \iff m(i) \neq m(j)$. Define now for all $1 \leq i \leq n^*$:

$$\begin{aligned}
&T^{\downarrow}(i) = \{x \in X^{\downarrow} : x\mathbf{R}^c m(i)\},\\
&T^{\uparrow}(i) = \{x \in X^{\uparrow} : \exists y \in X^{\downarrow}, y\mathbf{R}^c m(i), y\mathbf{R}^c x\}, \text{and}\\
&T(i) = T^{\uparrow}(i) \cup T^{\downarrow}(i).
\end{aligned}$$

Define finally:

$$T(0) \equiv X \backslash \tilde{X} = \bigcap_{y \in X} \bigcap_{t \in X} \{x \in X : \neg[x\mathbf{R}^c y], \neg[t\mathbf{R}^c x]\}.$$

These will define the types. We denote $\mathcal{T} = \{T(i) : 0 \leq i \leq n^*\}$ the collection of types.

Lemma 6. *$\mathcal{T}$ forms a partition of X.*

Proof. Given the definition of $T(0)$, it is sufficient to show that the collection $\{T(i) : 1 \leq i \leq n^*\}$ partitions $\tilde{X}$.

We first show that $\tilde{X} = \bigcup_{1 \leq i \leq n^*} T(i)$. Note that for all $1 \leq i \leq n^*$, if $x \in T(i)$, then there exists y such that $x\mathbf{R}^c y$ or $y\mathbf{R}^c x$, so that $x \in \tilde{X}$. Hence, $\bigcup_{i \leq n} T(i) \subseteq \tilde{X}$. Similarly, if $x \in \tilde{X}$, then either $x \in X^{\downarrow}$ or $x \in X^{\uparrow}$. If $x \in X^{\downarrow}$, then $x\mathbf{R}^c y$ for some $y \in X$ and by (11), $x\mathbf{R}^c m(x)$, i.e. $x \in T^{\downarrow}(i)$ for some $1 \leq i \leq n^*$. If $x \in X^{\uparrow}$, then $y\mathbf{R}^c x$ for some $y \in X$ and $\neg[x\mathbf{R}^c z]$ for all $z \in X$. But then $x \in R^{\uparrow}(y)$ and (11) implies that $y\mathbf{R}^c m(y) = m(i)$ for some $1 \leq i \leq n^*$. Therefore $x \in T^{\uparrow}(i)$. Hence, in both cases, $x \in \bigcup_{1 \leq i \leq n^*} T(i)$.

We now assume that for some $1 \leq i, j \leq n^*$, $x \in T(i) \cap T(j)$, and show that this implies $i = j$.

Case 1: Assume $x \in T^{\downarrow}(i)$. Then, because $X^{\uparrow}$ and $X^{\downarrow}$ are disjoint, x necessarily belongs to $T^{\downarrow}(j)$. $x \in T^{\downarrow}(i)$ means that $x\mathbf{R}^c m(i)$. By definition of the $m(i)$'s, there exists y such that $m(y) = m(i)$. Applying lemma 5, we conclude that $m(x) = m(y) = m(i)$. Similarly, we prove that $m(x) = m(j)$. Hence $m(i) = m(j)$ and therefore $i = j$.

Case 2: Assume $x \in T^{\uparrow}(i)$. Then, because $X^{\uparrow}$ and $X^{\downarrow}$ are disjoint, $x \in T^{\uparrow}(j)$. Hence, there exists $y_i, y_j \in X^{\downarrow}$ such that $y_i\mathbf{R}^c m(i)$, $y_j\mathbf{R}^c m(j)$, $y_i\mathbf{R}^c x$, and $y_j\mathbf{R}^c x$. Hence, $x \in R^{\uparrow}(y_i) \cap R^{\uparrow}(y_j)$, which by lemma 5, implies that $m(y_i) = m(y_j)$. Using the same argument as in *case 1*, we conclude that $m(i) = m(y_i) = m(y_j) = m(j)$, which means that $i = j$. □

Note that, given lemma 6, $T(x)$ is well defined as the type of the option $x \in X$, i.e. $T(x) = T(i) \iff x \in T(i)$.

Lemma 7. *For all $1 \leq i \leq n^*$, $x \in T^{\downarrow}(i)$ and $y \in T^{\uparrow}(i)$, $x \prec y$.*

Proof. If $y = m(i)$ this follows directly. If $y \neq m(i)$, there exists $z \in X$ such that $z\mathbf{R}^c y$ and $z\mathbf{R}^c m(i)$. Hence, $y, m(i) \in R^{\uparrow}(z)$ and $y \succ m(i)$. Moreover, $m(i) \in X^{\uparrow}$ so that $\neg[m(i)\mathbf{R}^c t]$ for any t. Since $x\mathbf{R}^c m(i)$, **R-Con**(ii) implies that $y \succ x$. □

Lemma 8. *For any $x, y \in X^{\downarrow}$, if $x\mathbf{R}^c y$, then $x\mathbf{P}^c y$.*

Proof. Let z, t such that $y\mathbf{R}^c_t z$. By**R-Tran**, it must be that $xR^c z$. Suppose first that $x \succ t$, then $x \succ t \succ z \succ x$. Then, if $zR^c t$, **R-Tran** implies that $yR^c t$, a contradiction. If $tR^c x$, **R-Tran** implies that $tR^c z$, a contradiction. Hence it must be that $xR^c_t z$, which means that $xP^c y$.

Suppose now that $t \succ x$. Let t' such that $xR^c_{t'} y$. Hence $x \succ t' \succ y$.

Case 1: $t' \succ t$. Then $x \succ t' \succ t \succ x$. First note that $t' \succ y \succ t$ and $\neg[tR^c y]$ (as otherwise this would imply $tR^c z$), $\neg[yR^c t']$ (as otherwise this would imply $xR^c t'$). Hence **R-NTran**(i) implies that $\neg[tR^c t']$.

1.1: $t'R^c x$. This implies that $t'R^c z$. Hence $y \succ x \succ t'$. Given that in addition $yR^c z$ we can apply **R-Mon**(i) to conclude that $xP^c y$, which contradicts that $t \succ x$.

1.2: xR^ct. Given that $t \succ y$, this means that by lemma 7, $t \in T^{\downarrow}(y)$ (given that $y \in T^{\downarrow}(y)$). Given that both xR^cz and yR^cz, we can apply **R-Mon**(i) to conlcude that tP^cy, which means that tR^cz a contradiction.

Case 2: $t \succ t'$. Then $y \succ t \succ t' \succ y$. First note that $t \succ x \succ t'$ and $\neg[t'R^cx]$ (as otherwise this would imply $t'R^cy$), $\neg[xR^ct]$ (as otherwise this would imply xP^cy as in case 1.2). Hence **R-NTran**(i) implies that $\neg[tR^ct']$.

2.1: yR^ct'. This implies that xR^ct', a contradiction.

2.2: $t \succ y$. This implies that tR^cz, again a contradiction.

Hence it must be that $x \succ t$ and therefore xP^cy. □

We now prove that $\succ$ is transitive on every type $T(i)$.

Lemma 9. *For all $0 \leq i \leq n^*$, the relation $\succ$ is transitive on $T(i)$.*

Proof. That $\succ$ is transitive on $T(0)$ is a direct consequence of lemma 3. We now focus on $1 \leq i \leq n^*$. We first show that $\succ$ is transitive on $T^{\downarrow}(i)$. Let $x, y, z \in T^{\downarrow}(i)$ such that $x \succ y \succ z$. Assume by contradiction that $z \succ x$. Suppose (w.l.o.g) that $x = c\{x, y, z\}$. In this case, $z\mathbf{R}^cy$. Given that $x\mathbf{R}^cm(i)$, $z\mathbf{R}^cm(i)$, and $x \succ y \succ z$, **R-Mon**(i) entails $y\mathbf{P^c}x$. But since $y, z \in X^{\downarrow}$, $z\mathbf{R}^cy$ implies $z\mathbf{P}^cy$, by lemma 8. Hence, by the transitivity of P^c (by definition), $z\mathbf{P}^cx$, which contradicts $z \succ x$.

Finally, we prove that $\succ$ is transitive on each type. Let i and $x, y, z \in T(i)$ such that $x \succ y \succ z$. If $x \in T^{\downarrow}(i)$ then, according to the first part of the proof, $y \in T^{\downarrow}(i)$ and therefore similarly $z \in T^{\downarrow}(i)$. Similarly, if $z \in T^{\uparrow}(i)$, the first part of the proof implies that $y \in T^{\uparrow}(i)$, which in turn also triggers that $x \in T^{\uparrow}(i)$. In both cases, we proved that $\succ$ is transitive on $T^{\downarrow}(i)$ and on $T^{\uparrow}(i)$ (a consequence of lemma 3). The last case is if $x \in T^{\uparrow}(i)$ and $z \in T^{\downarrow}(i)$, but then $x \succ z$ follows from lemma 7. □

For any menu A we define:

$$d(A) \equiv \{x \mid x = \max(T(x) \cap A, \succ)\}.$$

Lemma 9 implies that $d(A)$ is well defined. Furthermore, a direct implication of lemma 3 is that $\succ$ is transitive on $d(A)$. Hence we can state the following lemma:

Lemma 10. *For any $A \in \mathcal{X}$,*

$$c(A) = \max(d(A), \succ) \tag{12}$$

Proof. For any menu A, denote $i(A) = \#\{i \mid T(i) \cap A \neq \emptyset\}$. We prove that for any $1 \leq n \leq n^* + 1$, for any A such that $i(A) = n$, (12) holds.

If $i(A) = 1$, the conclusion follows from lemma 9. Assume now that $i(A) = 2$. Let $x, y \in A$ be such that $T(x) \cap T(y) = \emptyset$, $x = \max(T(x) \cap A, \succ)$, $y = \max(T(y) \cap A, \succ)$, and $y \succ x$. Assume by contradiction that $y \neq c(A)$. By definition of y, $y \succ z$ for any $z \in T(y) \cap A$. Hence, there must exist $z \in T(x)$ such that $z \succ y$ and $y \neq c\{x, y, z\}$, otherwise **Exp** would imply that $y = c(A)$. This implies that $x \succ z \succ y \succ x$. Since $y \neq c\{x, y, z\}$, this is only possible if either $y\mathbf{R}^c z$ or $x\mathbf{R}^c y$, which in any case contradicts that $x, z \notin T(y)$ (given that, according to lemma 6, types partition X). Hence we conclude that $y = c(A)$.

Then fix $3 \leq n \leq n^* + 1$ and let A a menu such that $i(A) = n$. We denote $y = \max(d(A), \succ)$. Given the preceding proof for any menu A' such that $i(A') = 2$, for any $z \in A$, $y = c\Big(\big(T(y) \cup T(z)\big) \cap A \Big)$. This implies by **Exp** that $y = c(A)$. $\square$

Lemma 11. *For any $x, y, z \in X$, if $xR^c yP^c z$, then $xR^c z$.*

Proof. Let $x, y, z \in X$ such that $xR^c yP^c z$. Let t such that $xR^c_t y$. A consequence of lemma 9 is that $t \notin T(x) = T(y) = T(z) \equiv T$. If $t \succ z$ then it suffices to prove that $xR^c z$. Suppose on the contrary that $z \succ t$. We want to show that there exists u such that $zR^c_u t$, contradicting $yP^c z$. Let u such that $zR^c u$, hence $u \in T$ and $u \succ z$. If $t \succ u$, then it suffices to prove that $zR^c_t u$. Suppose by contradiction that for any such u, $u \succ t$. Then there exists t' such that $zR^c_{t'} u$, in which case $yR^c_{t'} u$ and thus $t' \succ u \succ t \succ y \succ t'$. Suppose first that $t' \succ t$. Then $t' \succ t \succ y \succ t'$. Because $t, t' \notin T$, this implies that $tR^c t'$. But at the same time, given that $t' \succ u \succ t$, $t' \succ t$, $\neg[tR^c u]$ and $\neg[uR^c t']$, we can apply **R-Ntran**(i) to conclude that $\neg[tR^c t']$, a contradiction. An analogous reasoning applies for the other case $t \succ t'$ to obtain a contradiction. This ends the proof. $\square$

For all $0 \leq i \leq n^*$, define $\trianglerighteq_i$ on $T(i)$, for any $x, y \in T(i)$, $x \trianglerighteq_i y$ if one of the following cases is satisfied:

1. $xR^c y$;
2. $x \succ y \wedge \neg[yR^c x]$;
3. $xP^c y \wedge \neg[xR^c y]$.

Denote $\triangleright_i$ the asymmetric part of $\trianglerighteq_i$.

Lemma 12. *For all $0 \leq i \leq n^*$, $\trianglerighteq_i$ is complete and transitive.*

Proof. Note that by definition $\trianglerighteq_0$ is simply equal to $\succ$ on $T(0)$, hence it is complete and transitive. Let, $1 \leq i \leq n^*$ and $x, y, z \in T(i)$ such that $x \trianglerighteq_i y \trianglerighteq_i z$. We detail all the cases.

1. $x \succ y$. Together with $x \trianglerighteq_i y$, this implies that $\neg[yR^c x]$.

    1.1 $\neg[yP^c x]$.

    1.1.1 $y \succ z$. This implies that $x \succ z$ and $\neg[zR^c y]$, which in turn implies that $\neg[zR^c x]$ by **R-NTran**(i). Which implies that $x \trianglerighteq_i z$.

    1.1.2 $z \succ y$.

    1.1.2.1 $yR^c z$.

    i. $x \succ z$. $zR^c x \implies yR^c x$, a contradiction, hence $\neg[zR^c x]$ and $x \trianglerighteq_i z$.

    ii. $z \succ x$. Given that $\neg[yR^c x]$ and $yR^c z$, **R-NTran**(i) implies $xR^c z$, and therefore $x \trianglerighteq_i z$.

    1.1.2.2 $\neg[yR^c z]$. This implies that $yP^c z$.

    i. $x \succ z$. $zR^c x$ and $yP^c z$ would imply that $yR^c z$ (lemma 11), a contradiction. Hence $\neg[zR^c x]$ and $x \trianglerighteq_i z$.

ii. $z \succ x$. By **R-NTran**(ii), yP^cz and $\neg[yP^cx]$ imply that xP^cz, and thus $x \trianglerighteq_i z$.

1.2 yP^cx.

1.2.1 $y \succ z$. This implies that $\neg[zR^cy]$ and $x \succ z$. Then, $\neg[yR^cx] \implies \neg[zR^cx]$. Hence $x \trianglerighteq_i z$.

1.2.2 $z \succ y$.

1.2.2.1 yR^cz.

i. $x \succ z$. zR^cx would imply yR^cx, a contradiction. Hence $\neg[zR^cx]$ and $x \trianglerighteq_i z$.

ii. $z \succ x$. If $\neg[x\mathbf{R}^cz]$, then $\neg[y\mathbf{R}^cx]$ implies that $\neg[y\mathbf{R}^cz]$, a contradiction. Hence, $x\mathbf{R}^cz$ and $x \trianglerighteq_i z$.

1.2.2.2 $\neg[y\mathbf{R}^cz]$. This implies that $y\mathbf{P}^cz$.

i. $z \succ x$. By **R-Mon**(ii), yP^cz and yP^cx imply that xP^cz. Hence, $x \trianglerighteq_i z$.

ii. $x \succ z$. zR^cx together with yP^cz would imply yR^cx, a contradiction. Hence $\neg[zR^cx]$ and thus $x \trianglerighteq_i z$.

2. $y \succ x$.

2.1 xR^cy.

2.1.1 $z \succ y$. This implies that $z \succ x$.

2.1.1.1 yR^cz. Then **R-Tran** implies that xR^cz, i.e., $x \trianglerighteq_i z$.

2.1.1.2 $\neg[yR^cz]$. This implies that yP^cz. $xR^cyP^cz \implies xR^cz$. Hence $x \trianglerighteq_i z$.

2.1.2 $y \succ z$. This implies that $\neg[zR^cy]$.

2.1.2.1 $z \succ x$. If $\neg[xR^cz]$, then $\neg[zR^cy]$ implies that $\neg[xR^cy]$ a contradiction. Hence xR^cz and $x \trianglerighteq_i z$.

2.1.2.2 $x \succ z$. zR^cx would imply zR^cy, a contradiction, hence $x \trianglerighteq_i z$.

2.2 $\neg[xR^cy]$. This implies that xP^cy.

2.2.1 $z \succ y$. This implies that $z \succ x$.

2.2.1.1 yR^cz. Then xP^cy implies that xR^cz, i.e., $x \trianglerighteq_i z$.

2.2.1.2 $\neg[yR^cz]$. This implies that yP^cz. $xP^cyP^cz \implies xP^cz$. Hence $x \trianglerighteq_i z$.

2.2.2 $y \succ z$. This implies that $\neg[zR^cy]$.

2.2.2.1 $z \succ x$. xP^cy implies that $y \in T^{\downarrow}(i)$, and so $y \succ z$ implies that $z \in T^{\downarrow}(i)$ (lemma 7). So let $u \in T(i)$ such that zR^cu; $u \succ z$ and thus $u \succ x$. Let $t \notin T(i)$ such that $t = c\{u, z, t\}$ and $z \succ t$. Then because $\neg[zR^cy]$, it must be that $y \succ t$, which in turn implies that $x \succ t$ (as if $t \succ x$, we would conclude from **R-NTran**, $\neg[xP^ct]$ and $\neg[tP^cy]$ that $\neg[xP^cy]$). But from lemma 10, we get that $t = c\{x, u, t\}$, hence xP^cz and $x \trianglerighteq_i z$.

2.2.2.2 $x \succ z$. zR^cx would imply zR^cy, a contradiction, hence $\neg[zR^cx]$ and $x \trianglerighteq_i z$.

□

Lemma 13. *For any* $1 \leq i \leq n^\star$, $x, y, z \in T^{\downarrow}(i)$ *such that* $x \prec y \prec z$*:*

1. *if* $x \triangleright_i y$*, then* $y \trianglerighteq_i z$*;*
2. *if* $z \triangleright_i y$*, then* $y \trianglerighteq_i x$*.*

Proof. 1. $x \triangleright_i y$ implies that xR^cy. Given that $x, z \in T^{\downarrow}(i)$, $xR^cm(i)$ and $zR^cm(i)$. We can thus apply **R-Mon**(i) to conclude that yP^cz and thus $y \trianglerighteq_i z$.
2. $z \triangleright_i$ implies that $z \succ y \wedge \neg[yP^cz]$. Suppose bvy contradiction that $x \triangleright_i y$, i.e., xR^cy. Then, similarly as in case 1, **R-Mon**(i) implies that yP^cz, a contradiction. Therefore, $\neg[xR^cy]$ and thus $y \trianglerighteq_i x$. □

Denote $\tilde{\trianglerighteq} = \bigcup_i \trianglerighteq_i$. Let $\trianglerighteq$ be the relation on X defined by:

$$\forall x, y \in X, x \trianglerighteq y \iff \begin{cases} x \,\tilde{\trianglerighteq}\, y & \text{if } x \in T(y) \\ x \succ y & \text{if } x \notin T(y) \end{cases}$$

Denote $\triangleright$ the asymmetric part of $\trianglerighteq$.

Lemma 14. *The relation $\trianglerighteq$ is a complete preorder.*

Proof. We only need to prove the transitivity. Let $x, y, z \in X$ such that $x \trianglerighteq y \trianglerighteq z$. If there exists i such that $x, y, z \in T(i)$, then this follows from lemma 12. If $T(x) \cap T(y) = T(x) \cap T(z) = T(y) \cap T(z) = \emptyset$, then this follows from lemma 3.

Suppose we are in the case $T(x) = T(y) \neq T(z)$. Note that this implies that $y \succ z$. Suppose by contradiction that $z \succ x$. If $x \succ y$, this would imply that $yR^c x$, which contradicts $x \trianglerighteq y$, thus $y \succ x$. Given that $\neg[xR^c z]$ and $\neg[zR^c y]$, **R-NTran**(i) implies that $\neg[xR^c y]$. Similarly $\neg[xP^c y]$. Hence $y \triangleright x$, a contradiction. Hence $x \succ z$ and thus $x \triangleright z$. We deal with the case $T(y) = T(z) \neq T(x)$ similarly.

Suppose finally that we are in the case $T(x) = T(z) \neq T(y)$. Note that this implies that $x \succ y \succ z$. If $z \succ x$, then $xR^c z$ and thus $x \triangleright z$. Suppose on the contrary that $x \succ z$. Then **R-NTran** implies that $\neg[zR^c x] \wedge \neg[zP^c x]$, that is $x \triangleright z$. □

Now let $F = \bigcup_{1 \leq i \leq n^*} T^{\uparrow}(i) \cup T(0)$. Given that $\succ$ is transitive on F (lemma 3), there exists a function $w : F \longrightarrow \mathbb{R}$ that represents $\succ$ on F. Furthermore, for every $i = 0, \ldots, n^*$, $\succ$ is transitive on $T(i)$ (lemma 9), hence there exists a function $u_i : T(i) \longrightarrow \mathbb{R}$ representing $\succ$ on $T(i)$, and such that $u_T(x) = w(x)$ for every $x \in F$. We now define the function $u : X \longrightarrow \mathbb{R}$ such that for every i, $x \in T(i)$, $u(x) = u_i(x)$. We clearly have, for any menu A,

$$d(A) = \bigcup_{T \in \mathcal{T}} \arg \max_{x \in T \cap A} u(x).$$

Given lemma 14, there exists $v : X \longrightarrow \mathbb{R}$ that represents $\trianglerighteq$. Note that $\trianglerighteq \cap F^2 = \succ \cap F^2$, hence we can force that

$$v|_F = u|_F. \tag{13}$$

Lemma 15. *For any menu A,*

$$c(A) = \arg \max_{x \in d(A)} v(x).$$

Proof. For any $x, y \in d(A)$, $T(x) \neq T(y)$, so $\trianglerighteq \cap\, d(A)^2 = \succ \cap\, d(A)^2$. Therefore:

$$\arg\max_{x \in d(A)} v(x) = \max\big(d(A), \trianglerighteq\big) = \max\big(d(A), \succ\big) \underbrace{=}_{lemma\ 10} c(A).$$

□

Let $T \in \mathcal{T}$, so $T = T(i)$ for some $0 \leq i \leq n^\star$. Set $\lambda_T \equiv u(m(i))$.

To complete the proof of the theorem, we check that the tuple $\langle \mathcal{T}, \{\lambda_T\}, u, v\rangle$ so defined is an RS-structure that rationalizes the choice function c. Note first that by definition the order induced by u on each $T \in \mathcal{T}$ is $\succ \cap\, T^2$, which is a linear order (lemma 9). Lemma 15 shows that point (i) in definition 3 is satisfied. (13) show that point (ii) is satisfied. Finally, lemma 13 shows that point (iii) is satisfied. □

C Proof of Theorem 2

Proof. The necessity part of the theorem is left to the readers. We only prove the sufficiency.

(a) We first show that for any A, B such that $A \subseteq T$ and $B \subseteq T'$ for some $T, T' \in \mathcal{T}$, $A \succ B \iff A \cap F \neq \emptyset = B \cap F$. If $T = T'$, this is simply a consequence of part *(i)* of R-Dominance (RD).

Suppose now that $T \neq T'$. Let denote $A' = A \setminus F = \{a_1, \ldots, a_n\}$ and $B' = B \setminus F = \{b_1, \ldots, b_l\}$ and suppose that both are non-empty. By RD, $\{a_1\} \sim A'$, because both are richer than each other. Similarly $\{b_1\} \sim B'$. Furthermore, RD *(ii)* implies that $\{a_1\} \sim \{b_1\}$; hence, by transitivity, $A' \sim B'$.

Let denote $A'' = A \setminus A'$ and $B'' = B \setminus B'$. If $A'' = B'' = \emptyset$, we conclude from the previous argument that $A \sim B$. Suppose that $A'' \neq \emptyset = B''$, so $B = B'$. By a simple application of RD *(i)*, A is strictly richer than A', so $A \succ A'$, and by transitivity, $A \succ B$.

Assume now that $B'' \neq \emptyset$. By a similar reasoning as for A' and B', one can easily show that $A'' \sim B''$. If $B' = \emptyset$, then $B = B''$, hence $A \sim B$. If $B' \neq \emptyset$, note that $A' \cap A'' = B' \cap B'' = \emptyset$ and neither A' is richer than A'' nor B' is richer than

B''. Hence applying twice R-Composition (RC), we obtain that $A \sim B$.

(b) We next show that for any A, B, if $\#\Phi(A) = \#\Phi(B)$, then $A \sim B$. Denote $\Phi(A) = \{A_1, \ldots, A_n\}$ and $\Phi(B) = \{B_1, \ldots, B_n\}$. By (a), we know that for any i, $A_i \sim B_i$. Noting that $A_1 \cap A_2 = B_1 \cap B_2 = \emptyset$, and neither A_1 is richer than A_2 nor B_1 is richer than B_2, by applying twice RC, we get that $A_1 \cup A_2 \sim B_1 \cup B_2$. By reiterating the same argument, we obtain that $\bigcup_i A_i \sim \bigcup_i B_i$. Finally, note that A is richer that $\bigcup_i A_i$ and conversely $\bigcup_i A_i$ is richer than A, hence, by RD, $A \sim \bigcup_i A_i$; similarly $B \sim \bigcup_i B_i$. Therefore, by transitivity, we obtain that $A \sim B$.

(c) We finally prove that for any A, B, if $\#\Phi(A) > \#\Phi(B)$, then $A \succ B$. Denote $\Phi(A) = \{A_1, \ldots, A_n\}$ and $\Phi(B) = \{B_1, \ldots, B_k\}$, with $k < n$. By (b) $\bigcup_{i=1}^k A_i \sim B$. Furthermore, by RD, $\bigcup_{i=1}^n A_i \succ \bigcup_{i=1}^k A_i$. A similar argument as before shows that $A \sim \bigcup_{i=1}^n A_i$ and $B \sim \bigcup_{i=1}^k B_i$. Hence by transitivity $A \succ B$.

□

D Proofs of Section 5

Proof of Proposition 5. (i) Denote $u(\sigma^L)$ and $v(\sigma^{RR})$ the DM's anticipated utility from choosing respectively σ^L and σ^{RR} in the menu N:

$$u(\sigma^L) \leq v(\sigma^{RR}) \iff (1-p) + p\lambda - p(1-\lambda) \leq p + (1-p)\delta$$
$$\iff p \geq \frac{1-\delta}{3-2\lambda-\delta} = \frac{1/2}{5/2-2\lambda}$$

We define $p^\star := \frac{1/2}{5/2-2\lambda}$ and verify that $p^\star < 1/2$:

$$p^\star < 1/2 \iff \lambda < \frac{3}{4}$$

which is true by assumption.

(ii). We first compute the value $q^\star$ of the posterior such that for any $q \geq q^\star$, action r is preferred. $q^\star$ solves $(1-q) - q = q$, hence $q^\star = 1/3$.

Then we simply compare the posterior obtained after the realisation of a signal s^r from the news source σ^{RR} with $1/3$. The posterior is, $\frac{p}{p+(1-p)1/2}$, which is greater or equal than p. We are in the case where $p \geq p^\star$, hence it is sufficient to show that $p^\star \geq 1/3$: $p^\star \geq \frac{1}{3} \iff \lambda \geq \delta$ which is true by assumption. □

Proof of Lemma 1. The maximand of the program (5) is strictly concave and the set K_g is compact. Hence, C is well-defined (Weierstrass theorem) and is a choice function.

Now let us build the RS-structure $\langle \mathcal{T}, F, u, v\rangle$ that represents C. Given (6), $d^\star$ strictly increases with g if and only if $g > \hat{g}$. Let $g(t,d)$ be the g such that $t + gd^\beta = 1$.

Let us introduce the three following sets:

$$D^\uparrow = \bigcup_{\substack{t\in[0,1]\\ g>\hat{g}}} \{d \in [0,1] : (t,d) = C(K_g)\},$$

$$\forall d \in D^\uparrow, T(d) = \bigcup_{t\in[0,1]} \{(t,d)\},$$

$$T_0 = \bigcup_{\substack{d\notin D^\uparrow\\ t\in[0,1]}} \{(t,d)\}.$$

From these sets we can define the set of types and the freedom set

$$\mathcal{T} = \{T_0\} \cup \bigcup_{d\in D^\uparrow} \{T(d)\} \text{ and } F = T_0 \cup \bigcup_{d\in D^\uparrow} \{(t,d) \in T(d) : g(t,d) \leq \hat{g}\}$$

Now let us define u and v. For each (t,d) we posit $u(t,d) = t + P(d)\hat{V}$ and $v(t,d) = t + P(d)V\big(g(t,d)\big)$.

Given the uniqueness of $g(t,d)$ for each (t,d), v is well-defined. It can easily be shown that $\langle \mathcal{T}, F, u, v\rangle$ is an RS-structure. Consider the choice function C' which is the RSC defined on the compact subsets of $[0,1]^2$ and represented by $\langle \mathcal{T}, F, u, v\rangle$. We claim that for all g, $C(K_g) = C'(K_g)$. To check this claim let (t,d) and g such that $(t,d) = C(K_g)$ and (t',d') such that $(t',d') = C'(K_g)$. Note that $(t,d) = C(K_g)$ implies $g = g(t,d)$. Similarly, $(t',d') = C'(K_g)$ implies that $u(t',d') \geq u(t'',d')$ for all t'' such that $(t'',d') \in K_g$, that is, for all $t'' \leq t'$. Hence, $g = g(t',d')$.

Assume first that $g \leq \hat{g}$. Then, note that $(t', d') \in F$. Suppose that there exists $(t'', d'') \in K_g \setminus F$, this means that $g(t'', d'') > \hat{g}$, and hence there exists $t''' > t''$, such that $g(t''', d'') = g$, which implies $(t''', d'') \in K_g$ and $u(t''', d'') > u(t'', d'')$. Therefore, only elements in F can be considered for choices in K_g according to the choice procedure (3). Hence, both (t, d) and (t', d') are elements of $\arg\max u(K_g)$. Because the latter is a singleton, $(t, d) = (t', d')$.

Assume next that $g(t, d) > \hat{g}$. Suppose that $(t', d') \in F$, then this implies that $d' \notin D^{\uparrow}$, that is $d' \neq d$. Because, $g(t, d) = g(t', d')$, this in turn implies that $t' \neq t$. Furthermore, by definition, $(t, d) \notin F$, and from $(t', d') = C'(K_g)$, we deduce that $u(t', d') > v(t, d)$. We also know that $v(t', d') \geq u(t', d')$, so $v(t', d') > v(t, d)$, which contradicts that $(t, d) = C(K_g)$. Therefore, $(t', d') \notin F$. Because $(t, d) \notin F$, $(t', d') = C'(K_g)$ and $(t, d) = C(K_g)$ imply that $v(t', d') \geq v(t, d) = \max v(K_g)$. Therefore, $(t', d') \in \arg\max v(K_g)$, and given that this set is a singleton, this implies that $(t', d') = (t, d)$. □

Proof of Proposition 6. This is a straightforward consequence of (7). □

Proof of Proposition 7. Action a^L can only be implemented in the absence of both a^R and a^{LL}. In any case, if a^L is chosen in a menu M by the agent, it is chosen in both states L and R, which gives the principal the expected payoff:

$$(1-p)\pi_L(a^L) + p\pi_R(a^L). \tag{14}$$

Similarly, action a^R can only be implemented in the absence of a^{RR}, in which case it is chosen in both states L and R, giving the principal the expected payoff:

$$(1-p)\pi_L(a^R) + p\pi_R(a^R). \tag{15}$$

From this we can deduce the existence of $p_\star \in (0, 1)$ and $p^\star \in (0, 1)$ such that: for any $p < p_\star$, the principal strictly prefers a menu M (e.g. $\{a^L\}$) such that $a_\theta(M) = a^L$ for $\theta = L, R$; for any $p > p^\star$, the principal strictly prefers a menu M (e.g. $\{a^R\}$) such that $a_\theta(M) = a^R$ for $\theta = L, R$. Furthermore, there exists $\hat{p}$, the unique belief such that (14) = (15).

Only actions a^{LL} and a^{RR} can be simultaneously implemented respectively in state L and R. Given that $\pi_L(a^{LL}) > \pi_L(a^{RR})$ and $\pi_R(a^{RR}) > \pi_R(a^{LL})$, the principal will always prefer a menu implementing both actions (e.g. $\{a^{LL}, a^{RR}\}$) than a menu implementing only one of them. In this, the principal's expected payoff is:

$$(1-p)\pi_L(a^{LL}) + p\pi_R(a^{RR}). \tag{16}$$

Hence there exist $\underline{p}$ and $\bar{p}$ such that: (14) $\geq$ (16) if and only if $p \leq \underline{p}$; and (15) $\geq$ (16) if and only if $p \geq \bar{p}$.

The conclusions of the proposition follows easily from these observations.

□

E Comparisons of Welfare Criteria

By means of examples we show that neither the criterion $P^{\star}$ of Bernheim and Rangel (2009) nor the preference identified from *choice with limited attention* P^R (Masatlioglu, Nakajima and Ozbay, 2012) coincide with $\succ_c^w$.[32]

Example 1: $\succ_c^w \not\subset P^{\star}$. Let $z = c\{x, y, z\}$ and $x = c\{x, z\}$ and suppose that $u(z) \geq \lambda_{T(z)}$, $u(y) \geq \lambda_{T(y)}$. Then $z \succ_c^w x$ whilde $\neg[zP^{\star}x]$.

Example 2: $P^{\star} \not\subset \succ_c^w$. Consider x such that $u(x) < \lambda_{T(x)}$, $z \notin T(x)$ such that for every $y \in T(x)$, $v(y) > v(z)$. Then $xP^{\star}z$ while $\neg[x \succ_c^w z]$.

Example 3: $\succ_c^w \not\subset P^R \wedge P^r \not\subset \succ_c^w$. It is easy to see in the example given by Masatlioglu, Nakajima and Ozbay (2012) to show the difference between P^R and $P^{\star}$ (Example 1, pp. 2191-2192), that while they deduce xP^Ry we would concude that $y \succ_c^w$.

[32] $\succ_c^w = \succ_{\mathcal{S}}^w$ for any minimal RS-structure $\mathcal{S}$ that rationalizes c.

References

Abdelgadir, Aala, and Vasiliki Fouka. 2020. "Political Secularism and Muslim Integration in the West: Assessing the Effects of the French Headscarf Ban." *American Political Science Review*, 114(3): 707–723.

Adiwena, Bartolomeus Yofana, Monika Satyajati, and Widawati Hapsari. 2020. "Psychological Reactance and Beliefs in Conspiracy Theories During the Covid-19 Pandemic: Overview of the Extended Parallel Process Model (EPPM)." *Buletin Psikologi*, 28: 182.

Alesina, Alberto Francesco, and Bryony Reich. 2015. "Nation building." *Working Paper, Department of Economics, Harvard University*.

Alonso, Ricardo, and Niko Matouschek. 2008. "Optimal Delegation." *The Review of Economic Studies*, 75(1): 259–293.

Apesteguia, Jose, and Miguel A. Ballester. 2013. "Choice by sequential procedures." *Games and Economic Behavior*, 77(1): 90–99.

Apesteguia, Jose, and Miguel Ballester. 2015. "A Measure of Rationality and Welfare." *Journal of Political Economy*, 123(6): 1278 – 1310.

Arad, Ayala, and Ariel Rubinstein. 2018. "The People's Perspective on Libertarian-Paternalistic Policies." *The Journal of Law and Economics*, 61(2): 311–333.

Baujard, Antoinette. 2007. "Conceptions of freedom and ranking opportunity sets. A typology." *Homo Oeconomicus*, 2(24): 1–24.

Bernheim, B. Douglas, and Antonio Rangel. 2007. "Toward Choice-Theoretic Foundations for Behavioral Welfare Economics." *American Economic Review*, 97(2): 464–470.

Bernheim, B Douglas, and Antonio Rangel. 2008. "Choice-theoretic foundations for behavioral welfare economics." *The foundations of positive and normative economics: A handbook*, 155–192.

Bernheim, B. Douglas, and Antonio Rangel. 2009. "Beyond Revealed Preference: Choice-Theoretic Foundations for Behavioral Welfare Economics." *The Quarterly Journal of Economics*, 124(1): 51–104.

Bisin, Alberto, and Thierry Verdier. 2001. "The economics of cultural transmission and the dynamics of preferences." *Journal of Economic theory*, 97(2): 298–319.

Brehm, Jack W. 1966. *A theory of psychological reactance.* Academic Press.

Brehm, Sharon S, and Jack W Brehm. 2013. *Psychological reactance: A theory of freedom and control.* Academic Press.

Brock, Timothy C. 1968. "Implications of Commodity Theory for Value Change1." In *Psychological Foundations of Attitudes.* , ed. Anthony G. Greenwald, Timothy C. Brock and Thomas M. Ostrom, 243–275. Academic Press.

Bushman, Brad J. 2006. "Effects of Warning and Information Labels on Attraction to Television Violence in Viewers of Different Ages." *Journal of Applied Social Psychology*, 36(9): 2073–2078.

Chambers, Christopher P., and Takashi Hayashi. 2012. "Choice and individual welfare." *Journal of Economic Theory*, 147(5): 1818–1849.

Cherepanov, Vadim, Timothy Feddersen, and Alvaro Sandroni. 2013. "Rationalization." *Theoretical Economics*, 8(3): 775–800.

Chernoff, Herman. 1954. "Rational Selection of Decision Functions." *Econometrica*, 22: 422.

Che, Yeon-Koo, and Konrad Mierendorff. 2019. "Optimal Dynamic Allocation of Attention." *American Economic Review*, 109(8): 2993–3029.

Ehlers, Lars, and Yves Sprumont. 2008. "Weakened WARP and top-cycle choice rules." *Journal of Mathematical Economics*, 44(1): 87–94.

Fouka, Vasiliki. 2020. "Backlash: The unintended effects of language prohibition in US schools after World War I." *The Review of Economic Studies*, 87(1): 204–239.

Gosselt, Jordy, Menno De Jong, and Joris Van Hoof. 2012. "Effects of Media Ratings on Children and Adolescents: A Litmus Test of the Forbidden Fruit Effect." *Journal of Communication*, 62.

Grüne-Yanoff, Till. 2022. "What preferences for behavioral welfare economics?" *Journal of Economic Methodology*, 29(2): 153–165.

Hankin, Janet R., Ira J. Firestone, James J. Sloan, Joel Ager, Allen C. Goodman, Robert J. Sokol, and Susan S. Martier. 1993. "The Impact of the Alcohol Warning Label on Drinking during Pregnancy." *Journal of Public Policy & Marketing*, 12: 10 – 18.

Holmstrom, Bengt. 1980. "On The Theory of Delegation." Northwestern University, Center for Mathematical Studies in Economics and Management Science Discussion Papers.

Horan, Sean. 2016. "A simple model of two-stage choice." *Journal of Economic Theory*, 162: 372–406.

Hovland, Carl Iver, Irving Lester Janis, and Harold H Kelley. 1953. *Communication and persuasion; psychological studies of opinion change.* Yale University Press New Haven.

Jansen, Esther, Sandra Mulkens, and Anita Jansen. 2007. "Do not eat the red food!: Prohibition of snacks leads to their relatively higher consumption in children." *Appetite*, 49(3): 572–577.

Jansen, Esther, Sandra Mulkens, Yvette Emond, and Anita Jansen. 2008. "From the Garden of Eden to the land of plenty: Restriction of fruit and sweets intake leads to increased fruit and sweets consumption in children." *Appetite*, 51(3): 570–575.

Jonas, Eva, Verena Graupmann, Daniela Niesta Kayser, Mark Zanna, Eva Traut-Mattausch, and Dieter Frey. 2009. "Culture, self, and the emergence of reactance: Is there a "universal" freedom?" *Journal of Experimental Social Psychology*, 45(5): 1068–1080.

Jones, Peter, and Robert Sugden. 1982. "Evaluating choice." *International Review of Law and Economics*, 2(1): 47–65.

Kreps, David M. 1979. "A representation theorem for "preference for flexibility"." *Econometrica*, 565–577.

Levesque, Roger J. R. 2018. "Forbidden Fruit." *Encyclopedia of Adolescence*, , ed. Roger J. R. Levesque, 1453–1456. Springer International Publishing.

Lleras, Juan Sebastian, Yusufcan Masatlioglu, Daisuke Nakajima, and Erkut Y Ozbay. 2017. "When more is less: Limited consideration." *Journal of Economic Theory*, 170: 70–85.

Lynn, Michael. 1991. "Scarcity effects on value: A quantitative review of the commodity theory literature." *Psychology & Marketing*, 8(1): 43–57.

Manzini, Paola, and Marco Mariotti. 2007. "Sequentially rationalizable choice." *American Economic Review*, 97(5): 1824–1839.

Manzini, Paola, and Marco Mariotti. 2012. "Categorize then Choose: Boundedly Rational Choice and Welfare." *Journal of the European Economic Association*, 10(5): 1141–1165.

Masatlioglu, Yusufcan, Daisuke Nakajima, and Erkut Y Ozbay. 2012. "Revealed attention." *American Economic Review*, 102(5): 2183–2205.

Mazis, Michael B, Robert B Settle, and Dennis C Leslie. 1973. "Elimination of phosphate detergents and psychological reactance." *Journal of Marketing Research*, 10(4): 390–395.

Nehring, Klaus, and Clemens Puppe. 2002. "A theory of diversity." *Econometrica*, 70(3): 1155–1198.

Nyhan, Brendan, and Jason Reifler. 2010. "When corrections fail: The persistence of political misperceptions." *Political Behavior*, 32(2): 303–330.

Ok, Efe A, Pietro Ortoleva, and Gil Riella. 2015. "Revealed (p) reference theory." *American Economic Review*, 105(1): 299–321.

Pattanaik, Prasanta K., and Yongsheng Xu. 1990. "On Ranking Opportunity Sets in Terms of Freedom of Choice." *Recherches Économiques de Louvain / Louvain Economic Review*, 56(3/4): 383–390.

Pattanaik, Prasanta K, and Yongsheng Xu. 1998. "On preference and freedom." *Theory and Decision*, 44(2): 173–198.

Pattanaik, Prasanta K, and Yongsheng Xu. 2000. "On diversity and freedom of choice." *Mathematical Social Sciences*, 40(2): 123–130.

Pechmann, Cornelia, and Chuan-Fong Shih. 1999. "Smoking Scenes in Movies and Antismoking Advertisements before Movies: Effects on Youth." *Journal of Marketing*, 63(3): 1–13.

Ridout, Sarah. 2021. "Choosing for the Right Reasons." *Unpublished manuscript*.

Rosenberg, Benjamin, and Jason Siegel. 2018. "A 50-Year Review of Psychological Reactance Theory: Do Not Read This Article." *Motivation Science*, 4: 281–300.

Rubinstein, Ariel, and Yuval Salant. 2012. "Eliciting Welfare Preferences from Behavioural Data Sets." *The Review of Economic Studies*, 79(1): 375–387.

Samuelson, P. A. 1938. "A Note on the Pure Theory of Consumer's Behaviour." *Economica*, 5(17): 61–71.

Sen, Amartya. 1971. "Choice Functions and Revealed Preference." *Review of Economic Studies*, 38(July).

Sen, Amartya. 1993. "Markets and Freedoms: Achievements and Limitations of the Market Mechanism in Promoting Individual Freedoms." *Oxford Economic Papers*, 45(4): 519–541.

Sensenig, John, and Jack W Brehm. 1968. "Attitude change from an implied threat to attitudinal freedom." *Journal of Personality and Social Psychology*, 8(4p1): 324.

Sneegas, James E., and Tamyra A. Plank. 1998. "Gender differences in pre-adolescent reactance to age-categorized television advisory labels." *Journal of Broadcasting & Electronic Media*, 42(4): 423–434.

Varava, Kira A., and Brian L. Quick. 2015. "Adolescents and Movie Ratings: Is Psychological Reactance a Theoretical Explanation for the Forbidden Fruit Effect?" *Journal of Broadcasting & Electronic Media*, 59(1): 149–168.

Wood, Thomas, and Ethan Porter. 2019. "The elusive backfire effect: Mass attitudes' steadfast factual adherence." *Political Behavior*, 41(1): 135–163.