Political Science and Political Economy Working Paper

Department of Government

London School of Economics

No. 2/2014

Reason-Based Rationalization

Franz Dietrich (Paris School of Economics, CNRS, and University of East Anglia)

Christian List (LSE)

Reason-Based Rationalization

Franz Dietrich & Christian List*

June 2013 / January 2014

Abstract

"Reason-based rationalizations" explain an agent's choices by specifying which properties of the options or choice context he/she cares about (the "motivationally salient properties") and how he/she cares about these properties (the "fundamental preference relation"). We characterize the choice-behavioural implications of reason-based rationalizability and identify two kinds of context-dependent motivation in a reason-based agent: he/she may (i) care about different properties in different contexts and (ii) care not only about properties of the options, but also about properties relating to the context. Reason-based rationalizations can explain non-classical choice behaviour, including boundedly rational and sophisticated rational behaviour, and predict choices in unobserved contexts, an issue neglected in standard choice theory.

1 Introduction

The classical theory of individual choice faces many notorious problems. It is challenged by empirically well-established violations of rationality due to framing effects, menu-dependent choice, susceptibility to nudges, the use of heuristics, unawareness, and other related phenomena. For example, a redescription of the options can alter an agent's choice behaviour. Call this *the problem of bounded rationality*. The classical theory is also challenged by its inability to explain some intuitively rational but sophisticated forms of choice, such as choices based on norm-following or non-consequentialism. It does not distinguish these from ordinary rationality violations. For example, someone who never chooses the largest piece of cake offered to him (or her) for politeness and instead chooses the second largest violates the weak axiom of revealed preference and

*F. Dietrich, Paris School of Economics, CNRS, and University of East Anglia; C. List, London School of Economics. This work has been presented on numerous occasions, beginning with the LSE Choice Group workshop on "Rationalizability and Choice", July 2011. We thank the audiences at these occasions for helpful comments and suggestions.

thus counts as "irrational". Call this *the problem of sophisticated rationality*. We suggest that the theory's difficulty in addressing both problems stems from the lack of a model of how agents conceptualize options in any given choice context. When we provide such a model, a unified explanation of many of the challenging phenomena can be given.

Our basic idea (which can be viewed as developing a classical idea in consumer theory, e.g., Lancaster 1966) is the following. When an agent chooses between several options in some context, e.g., yoghurts in a supermarket, he (or she) conceptualizes each option not as a primitive object, but as a bundle of properties. Although each option can have many properties, the agent considers not all of them, but only a subset: the *motivationally salient* properties. In the supermarket, these may include whether the yoghurt is fruit-flavoured, low-fat, and free from artificial sweeteners, but exclude whether the yoghurt has an odd (as opposed to even) number of letters on its label (an irrelevant property) and whether it has been sustainably produced (a property ignored by many consumers). The agent then makes his choice on the basis of a *fundamental preference relation* over property bundles. He chooses one option over another, e.g., a low-fat cherry yoghurt over a full-fat, sugar-free vanilla one, if and only if his fundamental preference relation ranks the set of motivationally salient properties of the first option, say {low-fat, fruit-flavoured}, above the set of the second, say {full-fat, vanilla-flavoured, artificially sweetened}.

We call an agent's choice behaviour *reason-based rationalizable* if it can be explained in this way. More precisely, a *reason-based rationalization* explains the agent's choices by specifying (i) which properties he cares about in each choice context and (ii) how he cares about these properties. We formalize part (i) by a *motivational salience function*, which assigns to each context a set of motivationally salient properties, and part (ii) by a *fundamental preference relation* over property bundles.

Crucially, the motivationally salient properties may include not only (i) *option properties*, which options have independently of the choice context (and which are thus "intrinsic" to the options), but also (ii) *relational properties*, which options have relative to the context, and (iii) *context properties*, which are properties of the context alone. "Fruit-flavoured" and "low-fat" (in yoghurts) are option properties; they depend solely on the yoghurt. Whether a yoghurt is the only cherry yoghurt or the cheapest on display are relational properties; they depend also on the other available yoghurts. Examples of context properties are whether the available yoghurts include premium brands (this depends solely on the menu of options) and whether there is cheerful background music (this depends on features of the context over and above the menu).

Reason-based rationalizations can capture two kinds of context-dependence in an

agent's motivation. First, the context may affect which properties are motivationally salient, so that the agent cares about different properties in different contexts. We call this *context-variant motivation*. For example, some contexts make the agent diet-conscious, others not. Second, the motivationally salient properties may go beyond option properties and include relational or context properties, so that the agent cares about the context or about how the options relate to it. We call this *context-regarding motivation*. For example, the agent cares about whether the choice of an option is polite in the given context or whether there are luxury options available.

Many boundedly rational and sophisticated rational forms of choice can be subsumed under these two kinds of context-dependence. Arguably, bounded rationality, such as susceptibility to framing, nudging, or dynamic inconsistency, often involves context-variant motivation. Sophisticated rationality, such as norm-following or non-consequentialism, often involves context-regarding motivation. (Of course, we do not claim that context-variance is always boundedly rational or that context-regardingness is always sophisticated.)

Note that while we take agents to *conceptualize* options as bundles of motivationally salient properties, we could not simply *define* each option as a bundle of motivationally salient properties. Since an agent may conceptualize the same option in terms of different properties in different contexts, we cannot know the agent's motivationally salient properties *ex ante*; they can be inferred, at most, after observing the agent's choices (Bhattacharyya, Pattanaik, and Xu 2011 make a similar observation). Moreover, the same option can have different properties in different contexts when the properties are relational. The same piece of cake can be the second-largest in one context and the largest in another, and thus "politely choosable" in the former context, but not in the latter.

In Section 2, we introduce our framework and discuss some examples. In Section 3, we examine the choice-behavioural implications of the two kinds of context-dependence. In Section 4, we show how choice behaviour can reveal the motivational salience function and the fundamental preference relation. In Section 5, we discuss the prediction of choices in unobserved contexts, a topic neglected in standard choice theory. One of the messages of this paper is that psychological adequacy in the rationalization of choice matters greatly for the prediction of an agent's future choices.

To the best of our knowledge, our framework is novel. There is, of course, a growing body of works in decision theory offering non-standard approaches to rationalization (e.g., Suzumura and Xu 2001; Kalai, Rubinstein, and Spiegler 2002; Manzini and Mariotti 2007, 2012; Salant and Rubinstein 2008; Bernheim and Rangel 2009; Man-

dler, Manzini, and Mariotti 2012; Masatlioglu, Nakajima, and Ozbay 2012; Cherepanov, Feddersen, and Sandroni 2013), but none of them take the present approach. In the Appendix, we briefly discuss two conceptually related papers by Bossert and Suzumura (2009) and Bhattacharyya, Pattanaik, and Xu (2011) about the phenomenon of context-dependence. Among our contributions is a response to problems identified by them. Our paper also formalizes a distinction drawn by Rubinstein (2006) between "internal" and "external" reasons for choice, which parallels our distinction between context-regarding and context-unregarding motivation. More extensive reviews of the literature can be found in our philosophical papers on preference formation and preference change (Dietrich and List 2013a,b) and in the monograph by Bossert and Suzumura (2010).[1]

2 A general framework

2.1 Observable primitives

The observable primitives of our framework are the following:

- A non-empty set X of *options*. Typical elements are x, y, z, ...
- A non-empty set $\mathcal{K}$ of *contexts*. On the classical ("extensional") definition, each context $K \in \mathcal{K}$ is a non-empty set $K \subseteq X$ of feasible options, which the agent may choose from. On a more general ("non-extensional") definition, each context $K \in \mathcal{K}$ *induces* a non-empty feasible set $[K] \subseteq X$, but may carry additional information about the choice environment. Formally, K could be a pair (Y, λ) of a feasible set $Y(=[K])$ and an environmental parameter λ, representing a cue, default, room temperature, background music, or even a state of the agent such as "sober" or "drunk". (This resembles a "frame" in Salant and Rubinstein 2008 or "set of ancillary conditions" in Bernheim and Rangel 2009.) We simply write K for $[K]$, as it is always unambiguous whether K refers to the context broadly defined or to the feasible set $[K]$ (e.g., in "$x \in K$", K refers to $[K]$).
- A *choice function* $C : \mathcal{K} \to 2^X$, which assigns to each context $K \in \mathcal{K}$ a non-empty set of chosen options in K (i.e., $C(K) \subseteq K$).

[1] Our results here do not overlap with those in Dietrich and List (2013a,b), which did not address the rationalization of choice. The present paper is the first to discuss reason-based rationalizations of choice functions, to distinguish two forms of context-dependent choice, to treat motivationally salient properties not as primitives but as derivable from choice behaviour, and to offer a reason-based analysis of choice prediction. On the logic of preferences, property-based preferences, and preference or attitude change, see also Liu (2010), Osherson and Weinstein (2012), and Dietrich (2012). For further philosophical discussions supporting a "reason-based" perspective, see Pettit (1991) and Dietrich and List (2012).

2.2 Properties

A choice in context K can be viewed as a choice among pairs of the form (x, K), where $x \in K$. We call the elements of $X \times \mathcal{K}$ *option-context pairs*.[2] We define properties as features of option-context pairs. Formally, a *property* is an abstract object, P, that picks out a subset $[P] \subseteq X \times \mathcal{K}$ called its *extension*, consisting of all option-context pairs that "have" or "satisfy" the property (thus properties are binary). We assume that the extension of any property is distinct from $\varnothing$ and from $X \times \mathcal{K}$. This rules out properties that are never satisfied or always satisfied.

Our definition allows distinct properties to have the same extension. This is useful for capturing framing effects in which the description of a property matters. For example, the properties "80% fat-free" and "20% fat" (in foods) have the same extension but different descriptions and may sometimes prompt different responses. In many applications, however, it suffices to identify properties with their extensions.

We distinguish between three kinds of properties:

Option properties: These are properties whose possession by an option-context pair depends only on the option, not on the context. Examples are "fat-free" or "vanilla-flavoured" (in yoghurts). Formally, P is an *option property* if

$$(x, K) \in [P] \Leftrightarrow (x, K') \in [P] \text{ for all } x \in X \text{ and } K, K' \in \mathcal{K}.$$

Context properties: These are properties whose possession by an option-context pair depends only on the context, not on the option. Examples are "offering more than one feasible option", "offering a Rolls Royce among the feasible options", and – if contexts specify the choice environment over and above the feasible set – the time ("it's evening"), the temperature ("it's a hot day"), a default ("such-and-such is the status quo"), or some other frame. Formally, P is a *context property* if

$$(x, K) \in [P] \Leftrightarrow (x', K) \in [P] \text{ for all } x, x' \in X \text{ and } K \in \mathcal{K}.$$

Relational properties: These are properties whose possesion by an option-context pair depends on both the option and the context, capturing their relationship. Examples are "not being the largest piece of cake offered" and "being the most expensive car on the market". Formally, P is a *relational property* if it is neither an option property nor a context property.

[2] Note that some pairs (x, K) in $X \times \mathcal{K}$ are "infeasible" in the sense that $x \notin K$.

We call properties that are not option properties *context-regarding* and properties that are not context properties *option-regarding*. Relational properties are context-regarding *and* option-regarding.

2.3 An example

To illustrate how properties may determine an agent's choice, we give an example to which we will refer repeatedly. It concerns the choice of fruit at a dinner party, as in Sen's well-known example of a polite dinner-party guest. Let X contain different fruits: apples, bananas, chocolate-covered pears, and possibly others. Each kind of fruit comes in up to three sizes: big, medium, and small. A choice context is a non-empty feasible set $K \subseteq X$, consisting of fruits currently in the basket. The set of possible contexts is $\mathcal{K} = 2^X \setminus \{\varnothing\}$. We consider the following properties:

- "big", "medium", and "small": the option properties of being a big, medium, and small fruit, respectively;
- "chocolate-offering": the context property of offering at least one chocolate-covered fruit among the feasible options;
- "polite": the relational property of not being the last available fruit of its kind, i.e., not being the last apple in the basket, the last banana, and so on.

We describe four agents whose choice behaviour we will subsequently explain:

Bon-vivant Bonnie always chooses a largest available fruit. For any K, she chooses

$$C(K) = \{x \in K : x \text{ is largest in } K\},$$

where "medium" is larger than "small", and "big" is larger than both other sizes.

Polite Pauline politely avoids choosing the last available fruit of its kind and only secondarily cares about a fruit's size. For any K, she chooses

$$C(K) = \{x \in K : x \text{ is largest in } K^* \text{ if } K^* \neq \varnothing \text{ and largest in } K \text{ if } K^* = \varnothing\},$$

where K^* is the set of all fruits in K that are not the last available ones of their kind.

Chocoholic Coco picks any fruit indifferently when no chocolate-covered fruit is available, but otherwise chooses a largest available fruit, because the smell of chocolate makes him hungry. For any K, he chooses

$$C(K) = \begin{cases} K & \text{if } K\text{contains no chocolate-covered fruit,} \\ \{x \in K : x \text{ is largest in } K\} & \text{otherwise.} \end{cases}$$

Weak-willed William makes the same polite choices as Pauline when no chocolate-covered fruit is available, and the same "greedy" choices as Bonnie otherwise, as the smell of chocolate makes him lose his inhibitions. For any K, he chooses

$$C(K) = \begin{cases} \{x \in K : x \text{ is largest in } K^*\} & \text{if } \left[\begin{array}{c} K\text{contains no chocolate-covered fruit} \\ \text{and } K^* \neq \varnothing \end{array} \right], \\ \{x \in K : x \text{ is largest in } K\} & \text{otherwise,} \end{cases}$$

where K^* is again the set of fruits in K that are not the last available ones of their kind.

2.4 Reason-based models

To explain an agent's choices, we consider a set $\mathcal{P}$ of potentially relevant properties, called a *property system*. It contains the properties we have at our disposal for any rationalization. In our example, $\mathcal{P} = \{\text{big, medium, small, chocolate-offering, polite}\}$. The specification of $\mathcal{P}$ may depend on our explanatory goals. The slimmer $\mathcal{P}$ is, the fewer patterns of choice can be explained. The set $\mathcal{P}$ can be partitioned into the sets $\mathcal{P}_{\text{option}}$, $\mathcal{P}_{\text{context}}$, and $\mathcal{P}_{\text{relational}}$ of option properties, context properties, and relational properties, respectively. For any option x and any context K, we write

- $\mathcal{P}(x, K)$ for the set $\{P \in \mathcal{P} : (x, K) \in [P]\}$ of all properties of the pair (x, K),
- $\mathcal{P}(x) = \mathcal{P}(x, K) \cap \mathcal{P}_{\text{option}}$ for the set of option properties of x, and
- $\mathcal{P}(K) = \mathcal{P}(x, K) \cap \mathcal{P}_{\text{context}}$ for the set of context properties of K.

Each of these three sets is assumed to be finite (while X, $\mathcal{K}$, and $\mathcal{P}$ need not be finite). A subset of $\mathcal{P}$ is called a *property bundle*.

We define a *reason-based model* of an agent, $\mathcal{M}$, as a pair $(M, \geq)$ consisting of:

- A *motivational salience function* M (formally a function from $\mathcal{K}$ into $2^{\mathcal{P}}$), which assigns to each context $K \in \mathcal{K}$ a set $M(K)$ of *motivationally salient* properties in context K. We require that any contexts with the same context properties induce the

same motivationally salient properties, i.e., if $\mathcal{P}(K)=\mathcal{P}(K')$ then $M(K)=M(K')$. (So differences in motivation must stem from differences in context properties.)

- A *fundamental preference relation* $\geq$ over property bundles (formally a binary relation on $2^{\mathcal{P}}$, on which we initially impose no restrictions). We write $>$ and $\equiv$ for the strict and indifference relations induced by $\geq$.

Informally, M specifies which properties the agent cares about in each context, and $\geq$ specifies how he cares about these properties, by ranking different property bundles relative to one another.

The model $\mathcal{M}$ represents (i) how the agent conceptualizes options in each context, (ii) what his resulting preferences over the options are, and (iii) what choices he is disposed to make. Formally:

- Any option x is *conceptualized* in context K as the set of motivationally salient properties of (x, K), denoted $x_K = \mathcal{P}(x, K) \cap M(K)$.

- The agent's *preference relation* in context K is the binary relation $\succsim_K$ on X defined as follows:

$$x \succsim_K y \Leftrightarrow x_K \geq y_K \text{ for all } x, y \in X.$$

  We write $\succ_K$ and $\sim_K$ for the strict and indifference relations induced by $\succsim_K$.

- The agent's *choice dispositions* are given by the function $C^{\mathcal{M}} : \mathcal{K} \to 2^X$ which assigns to each context the set of most preferred feasible options in that context:

$$C^{\mathcal{M}}(K) = \{x \in K : x \succsim_K y \text{ for all } y \in K\}.$$

  This defines an *improper choice function* ("improper" because $C^{\mathcal{M}}(K)$ may be empty for some K if $\geq$ is not well-behaved).

A choice function $C : \mathcal{K} \to 2^X$ is *reason-based rationalizable* (relative to $\mathcal{P}$) if there exists a reason-based model $\mathcal{M}$ (relative to $\mathcal{P}$) such that $C = C^{\mathcal{M}}$. We then call $\mathcal{M}$ a *rationalization* of C.

The fact that reason-based rationalizations depend on the property system $\mathcal{P}$ does not render them *ad hoc*. To the contrary, by introducing properties we can express hypotheses about how an agent conceptualizes and individuates the options. Classical choice theory treats options as exogenously given. In effect, the classical specification of the options also encodes a hypothesis about how options are individuated, though less transparently so. Thus our framework allows us to make explicit an issue that is largely neglected in classical choice theory.

2.5 Revisiting the example

The four choice functions in our example are all reason-based rationalizable:

Bon-vivant Bonnie's choice function can be rationalized by defining the set of motivationally salient properties in any context K as

$$M(K) = \{\text{big, medium, small}\} \text{ (so } M \text{ is a constant function)},$$

and defining the fundamental preference relation $\geq$ such that the three singleton property bundles {big}, {medium}, and {small} stand in the linear order satisfying

$$\{\text{big}\} > \{\text{medium}\} > \{\text{small}\}.^{3}$$

For instance, in a context K that offers only a small apple a and a big banana b, Bonnie chooses the banana b. She conceptualizes the two fruits as

$$a_K = \mathcal{P}(a, K) \cap M(K) = \{\text{small}\},$$
$$b_K = \mathcal{P}(b, K) \cap M(K) = \{\text{big}\},$$

and $b_K \succsim_K a_K$ since $\{\text{big}\} > \{\text{small}\}$.

Polite Pauline's choice function can be rationalized by defining the set of motivationally salient properties in any context K as

$$M(K) = \{\text{big, medium, small, polite}\} \text{ (so, again, } M \text{ is a constant function)},$$

and defining the fundamental preference relation $\geq$ such that the property bundles {big, polite}, {medium, polite}, {small, polite}, {big}, {medium} and {small} stand in the linear order satisfying

$$\{\text{big, polite}\} > \{\text{medium, polite}\} > \{\text{small, polite}\} > \{\text{big}\} > \{\text{medium}\} > \{\text{small}\}.$$

[3] Formally, $\geq$ = {({big},{big}), ({big},{medium}), ({big},{small}), ({medium},{medium}), ({medium},{small}), ({small},{small})}.

For instance, if only two small apples a and a' and one big banana b are available in context K, Pauline chooses an apple. She conceptualizes the three fruits as

$$\begin{aligned} a_K &= \mathcal{P}(a, K) \cap M(K) = \{\text{small, polite}\}, \\ a'_K &= \mathcal{P}(a', K) \cap M(K) = \{\text{small, polite}\}, \\ b_K &= \mathcal{P}(b, K) \cap M(K) = \{\text{big}\}, \end{aligned}$$

where $a_K \sim_K a'_K \succsim_K b_K$ since $\{\text{small, polite}\} \equiv \{\text{small, polite}\} > \{\text{big}\}$.

Chocoholic Coco's choice function can be rationalized by defining the set of motivationally salient properties in any context K as

$$M(K) = \begin{cases} \varnothing & \text{if no chocolate-covered fruit is available in } K, \\ & \quad \text{i.e., chocolate-offering} \notin \mathcal{P}(K), \\ \{\text{big, medium,} & \text{if a chocolate-covered fruit is available in } K, \\ \text{small}\} & \quad \text{i.e., chocolate-offering} \in \mathcal{P}(K), \end{cases}$$

and defining the fundamental preference relation $\geq$ as in Bonnie's case, with the only additional stipulation that $\varnothing \equiv \varnothing$. For instance, in a context without a tempting chocolate-covered fruit, he picks any fruit indifferently, because he conceptualizes every fruit as the same empty property bundle $\varnothing$, where $\varnothing \equiv \varnothing$.

Weak-willed William's choice function can be rationalized by defining the set of motivationally salient properties in any context K as

$$M(K) = \begin{cases} \{\text{big, medium,} & \text{if no chocolate-covered fruit is available in } K, \\ \text{small, polite}\} & \quad \text{i.e., chocolate-offering} \notin \mathcal{P}(K), \\ \{\text{big, medium,} & \text{if a chocolate-covered fruit is available in } K, \\ \text{small}\} & \quad \text{i.e., chocolate-offering} \in \mathcal{P}(K), \end{cases}$$

and defining the fundamental preference relation $\geq$ as in Pauline's case. So, if context K offers only two small apples a and a' and one big banana b, then, undisturbed by any smell of chocolate, he conceptualizes these fruits as Pauline does and politely chooses a small apple. If a small chocolate-covered pear is added to the basket, he forgets about politeness and conceptualizes the fruits as Bonnie does, choosing the big banana.

2.6 Two kinds of context-dependent motivation

In our example, Polite Pauline and Chocoholic Coco are affected by the context in opposite ways. Pauline *cares about* the context, since the relational property "polite" is motivationally salient for her. Coco's set of motivationally salient properties *varies with* the context: different contexts make him care about different properties. We say that an agent's motivation, according to model $\mathcal{M} = (M, \geq)$, is

- *context-regarding* if the range of the motivational salience function M includes not only sets of option properties (i.e., $M(K)$ contains at least one context-regarding property for some $K \in \mathcal{K}$), and *context-unregarding* otherwise;
- *context-variant* if M is a non-constant function (i.e., $M(K)$ is not the same for all $K \in \mathcal{K}$), and *context-invariant* otherwise.

How do the two kinds of context-dependence affect the agent's conceptualization of the options in each context? Table 1 shows how the agent conceptualizes any option x in context K, depending on which of the two kinds of context-dependence are present.

		Context-variant motivation? Yes	Context-variant motivation? No
Context-regarding motivation? Yes		$x_K = \mathcal{P}(x, K) \cap M(K)$ (e.g., William)	$x_K = \mathcal{P}(x, K) \cap M$ (e.g., Pauline)
Context-regarding motivation? No		$x_K = \mathcal{P}(x) \cap M(K)$ (e.g., Coco)	$x_K = \mathcal{P}(x) \cap M$ (e.g., Bonnie)

Table 1: The agent's conceptualization of option x in context K

Note that, with both kinds of context-dependence permitted, option x is conceptualized in context K as $x_K = \mathcal{P}(x, K) \cap M(K)$, which may depend on the context in two places: (i) in the set of properties of the option-context pair (x, K) and (ii) in the set of motivationally salient properties in context K. If the agent's motivation is context-unregarding, $\mathcal{P}(x, K)$ can be replaced by $\mathcal{P}(x)$. Here, $M(K)$ contains only option properties, so that $\mathcal{P}(x, K) \cap M(K) = \mathcal{P}(x) \cap M(K)$. If the agent's motivation is context-invariant, $M(K)$ can be replaced by a single set M of motivationally salient properties. Here, M is a constant function, so that the first component of the reason-based model $(M, \geq)$ can be redefined as a fixed set M. In the case of no context-dependence, the agent's conceptualization of option x in any context K simplifies to $x_K = \mathcal{P}(x) \cap M$.

From a classical perspective, agents with context-invariant motivation seem more rational than agents whose motivation varies as a result of subtle environmental features like the smell of chocolate. Bonnie exemplifies the case of classical rationality:

context-invariant motivation and context-unregarding conceptualization of the options. Pauline displays sophisticated rational behaviour: she considers not only properties of the options, but also properties concerning the relationship between options and context, such as politeness. William tries to display the same sophisticated behaviour, but is susceptible to variations in motivation across different contexts. Coco, finally, focuses only on option properties, but, like William, lacks a stable motivation.

2.7 Some illustrative non-classical choice behaviours

To illustrate the generality of this framework, we briefly show how it can represent framing effects, choices by heuristics or checklists, and non-consequentialist choices.

Framing effects: Framing effects can be understood as special kinds of choice reversals. A *choice reversal* occurs when there are contexts K and K' and options x and y such that x is chosen over y in K and y is chosen over x in K', where at least one choice is strict. (Option x is *chosen weakly over* option y in context K if $x, y \in K$ and $x \in C(K)$; and *strictly* if, in addition, $y \notin C(K)$.) Choice reversals can have two sources, according to a reason-based rationalization of C. The source is *context-variance* if K and K' induce different sets of motivationally salient properties $M(K) \neq M(K')$ both of which contain only option properties. The source is *context-regardingness* if K and K' induce the same set $M(K) = M(K')$, but this set contains some relational or context properties that distinguish the choice between x and y in the two contexts. (There are also mixed cases.) In either case, the agent prefers x to y *as conceptualized in context* K, and y to x *as conceptualized in context* K', as illustrated in Figure 1. We may define a

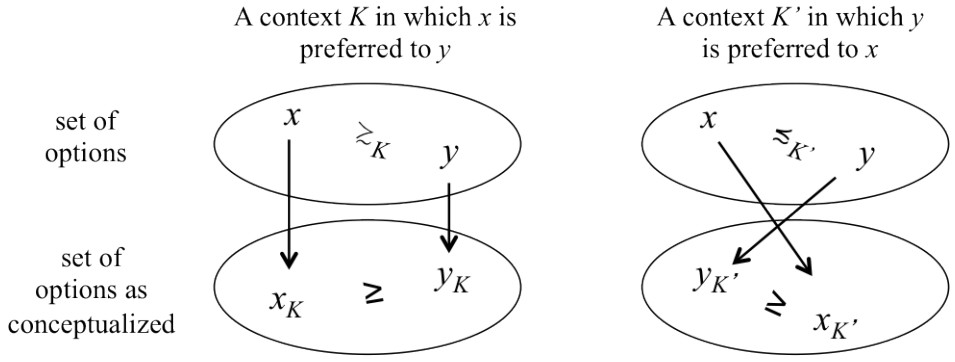

Figure 1: A choice reversal

framing effect as a choice reversal whose source is context-variance, and define the *frame* in each context K as the set of context properties $\mathcal{P}(K)$ "responsible" for $M(K)$. (In

Section 5, we introduce a notion of *causally relevant* context properties that could be used to refine this definition.) Whether a choice reversal counts as a framing effect then depends on the reason-based model by which we rationalize C. Note that, if K and K' offer the same feasible options, framing effects can occur only if contexts are defined non-extensionally, as consisting of both a feasible set and an environmental parameter (as in Salant and Rubinstein 2008); otherwise $M(K)$ and $M(K')$ would have to coincide. If K and K' offer different feasible options, framing effects can occur even when contexts are defined extensionally, provided they are distinguished by some context properties (such as "offering luxury goods") that lead to the difference between $M(K)$ and $M(K')$.

Checklists or "take-the-best" heuristics: Here, the agent considers a list of criteria by which the options can be distinguished and places the criteria in some order of importance. For any set of feasible options, the agent first compares the options in terms of the first criterion; if there are ties, he moves on to the second criterion; if there are still ties, he moves on to the third; and so on. Gigerenzer et al. (e.g., 2000) describe empirical examples of such choice procedures, and Mandler, Manzini, and Mariotti (e.g., 2012) offer a formal analysis (see also Liu 2010). In our framework, we can rationalize such choice behaviour by a reason-based model $(M, \geq)$ with a lexicographic fundamental preference relation $\geq$, where property bundles are ranked on the basis of some order of importance over properties. To illustrate, let P_1, P_2, P_3, ... denote the first, second, third, ..., properties in this order (assuming a finite $\mathcal{P}$). We can then define the fundamental preference relation $\geq$ as follows: for any property bundles S_1 and S_2, let $S_1 \geq S_2$ if and only if either $S_1 = S_2$ or there is some n such that (i) $P_n \in S_1$, (ii) $P_n \notin S_2$, and (iii) $S_1 \cap \{P_1, ..., P_{n-1}\} = S_2 \cap \{P_1, ..., P_{n-1}\}$. A lexicographic fundamental preference relation can be combined with either context-variant or context-invariant motivation, and with either context-regarding or context-unregarding motivation. This opens up greater generality than usually acknowledged.

Non-consequentialism: A non-consequentialist agent, in the most general sense, makes a choice in a given context not just on the basis of the chosen option itself (the "outcome"), but also on the basis of what the choice context is or how each option relates to that context (the "act of choosing the option"). Any context-regarding motivation can thus be viewed as a form of non-consequentialism. More narrowly, we may consider an agent who cares about whether each option is "permissible" or "norm-conforming" in a given context. The relevant criterion may be, for example, politeness, legality, or moral permissibility in the context. Let us introduce a relational property P such that any

option-context pair (x, K) satisfies P if and only if the choice of x is deemed permissible or norm-conforming in context K. If P is in every $M(K)$ and the fundamental preference relation ranks property bundles that include P above bundles that do not, the agent will always choose a permissible or norm-conforming option, unless no such option is feasible. Note that this could not generally be modelled without context-regarding motivation. For earlier discussions of non-consequentialist and "norm-conditional" choices, see, e.g., Suzumura and Xu (2001) and Bossert and Suzumura (2009).

3 Choice-behavioural implications

When does a choice function $C : \mathcal{K} \to 2^X$ have a reason-based rationalization? We first give necessary and sufficient conditions for reason-based rationalizability without any restriction, permitting both context-variant and context-regarding motivation. We then characterize the opposite case, without any context-dependence. Finally, we address the two intermediate cases, where rationalizability is restricted to either context-invariant or context-unregarding motivation but not both. The reader may skip this section if he or she is interested primarily in constructing reason-based models from observed choices (Section 4) or in predicting choices in novel contexts (Section 5).

3.1 Reason-based rationalizability without any restriction

We begin by stating two axioms which, together, imply that choice is based on properties. The first is an "intra-context" axiom. It states that the agent's choice in any context does not distinguish between options that have the same properties in that context:

Axiom 1 For all contexts $K \in \mathcal{K}$ and all options $x, y \in K$, if $\mathcal{P}(x, K) = \mathcal{P}(y, K)$, then $x \in C(K) \Leftrightarrow y \in C(K)$.

The second axiom is an "inter-context" axiom. It states that if two contexts offer the same feasible property bundles, the agent chooses options *instantiating the same property bundles* in those contexts:

Axiom 2 For all contexts $K, K' \in \mathcal{K}$, if $\{\mathcal{P}(x, K) : x \in K\} = \{\mathcal{P}(x, K') : x \in K'\}$, then $\{\mathcal{P}(x, K) : x \in C(K)\} = \{\mathcal{P}(x, K') : x \in C(K')\}$.

This is weaker than requiring that the *same options* be chosen in those contexts. The axiom further requires no relationship between the choices in contexts K and K'

with different context properties (i.e., $\mathcal{P}(K) \neq \mathcal{P}(K')$), since these automatically offer different feasible property bundles.

Axioms 1 and 2 do not by themselves imply any maximizing behaviour.[4] This gap is filled by our third axiom, a variant of Richter's (1971) axiom of "revelation coherence" (which, in turn, is a weakening of the weak axiom of revealed preference; see, e.g., Samuelson 1948). Unlike Richter, we formulate our axiom at the level of property bundles, not options. We adapt some revealed-preference terminology. For any property bundles S and S':

- S is *feasible* in context K if $S = \mathcal{P}(x, K)$ for some feasible option $x \in K$;
- S is *chosen* in context K if $S = \mathcal{P}(x, K)$ for some option $x \in C(K)$;
- S is *revealed weakly preferred* to S' (formally $S \succsim^C S'$) if, in some context, S is chosen while S' is feasible; S is *revealed strictly preferred* to S' if, in some context, S is chosen while S' is feasible and not chosen.[5]

Axiom 3 If a property bundle $S \subseteq \mathcal{P}$ is feasible in some context $K \in \mathcal{K}$ and is revealed weakly preferred to every feasible property bundle in context K, then S is chosen in context K.

Like Axiom 2, Axiom 3 is less restrictive than one might think. For the choices in context K to constrain those in context K', the two contexts must have the same context properties, i.e., $\mathcal{P}(K) = \mathcal{P}(K')$. Otherwise there will be no property bundles that are feasible in both K and K'. In fact:

Lemma 1 *Axiom 3 strengthens Axiom 2.*

Theorem 1 *A choice function C is reason-based rationalizable if and only if it satisfies Axioms 1 and 3 (and by implication 2).*[6]

[4] They are jointly equivalent to choice being rationalizable by a *generalized reason-based model*, defined by (i) a motivational salience function and (ii) a choice function defined on property bundles, not on options (which is more general than a fundamental preference relation $\geq$ over property bundles).

[5] One must not interpret the revealed-preference relation $\succsim^C$ as representing the agent's fundamental preferences. When the agent revealed-prefers bundle S to bundle S' by choosing S over S' in some context, only some *subsets* of S and S' are usually motivationally salient, and the fundamental preference is held between these, not between S and S'. In Section 4, we introduce a notion of revealed *fundamental* preference. The revealed-preference relation $\succsim^C$ between property bundles induces a context-variant revealed-preference relation $\succsim^C_K$ between options: option x is *revealed weakly preferred* to option y in context K ($x \succsim^C_K y$) if and only if $\mathcal{P}(x, K) \succsim^C \mathcal{P}(y, K)$. In classical choice theory, without the resources of properties, it is hard to define an interesting notion of context-variant revealed preference. Classical revealed preferences are context-invariant and fail to rationalize many observable choice behaviours.

[6] Axioms 1 and 3 are jointly equivalent to the requirement that, for every $K \in \mathcal{K}$ and every $x \in K$, if $\mathcal{P}(x, K)$ is revealed weakly preferred to $\mathcal{P}(y, K)$ for every $y \in K$, then $x \in C(K)$.

This result, like all subsequent results, holds for each property system $\mathcal{P}$. We can thus test for rationalizability in different property systems, e.g., by asking: Is the agent's choice between cars rationalizable in a system of colour-related properties? In a system of prestige-related properties? In a system of prestige- and price-related properties?[7]

Reason-based rationalizations need not be unique. For a given choice function C, there may exist more than one reason-based model $\mathcal{M}$ such that $C = C^{\mathcal{M}}$. Different rationalizations are far from equivalent, as discussed in detail later. They may lead to different predictions for novel choice contexts outside the set $\mathcal{K}$ of "observed" contexts, as shown in Section 5. We now reduce and later (in Section 4) eliminate the non-uniqueness of $\mathcal{M}$, by imposing additional restrictions on the admissible reason-based models.

3.2 Reason-based rationalizability without any context-dependence

While we have so far allowed rationalizations to display both kinds of context-dependence, we now consider the opposite, limiting case with no context-dependence at all. Consider the following variants of Axioms 1 and 2, obtained by referring only to context-unregarding properties:

Axiom 1* For all contexts $K \in \mathcal{K}$ and all options $x, y \in K$, if $\mathcal{P}(x) = \mathcal{P}(y)$, then $x \in C(K) \Leftrightarrow y \in C(K)$.

Axiom 2* For all contexts $K, K' \in \mathcal{K}$, if $\{\mathcal{P}(x) : x \in K\} = \{\mathcal{P}(x) : x \in K'\}$, then $\{\mathcal{P}(x) : x \in C(K)\} = \{\mathcal{P}(x) : x \in C(K')\}$.

In our example, Bon-vivant Bonnie satisfies both axioms; Chocoholic Coco satisfies Axiom 1* but violates Axiom 2* (to see this, suppose K contains a chocolate-covered pear while K' does not); and Polite Pauline and Weak-willed William violate even Axiom 1* (they care about a relational property).

We also introduce an analogue of Axiom 3, namely Richter's (1971) original axiom of *revelation coherence*, extended to our framework where contexts (if defined non-extensionally) can be more general than feasible sets.

Axiom 3* For all contexts $K \in \mathcal{K}$ and any feasible option $x \in K$, if, for every option $y \in K$, there is a context $K' \in \mathcal{K}$ in which x is chosen weakly over y, then $x \in C(K)$.

[7] To make this explicit, we could restate Theorem 1 (and similarly other results) as follows: *For every property system $\mathcal{P}$, a choice function C is reason-based rationalizable in $\mathcal{P}$ if and only if it satisfies Axioms 1 and 3 (and thereby 2).*

To state our characterization of reason-based rationalizability without any context-dependence, call the set of contexts $\mathcal{K}$ *closed under cloning* if $\mathcal{K}$ is closed under transforming any context by adding "clones" of feasible options; formally, whenever a context $K \in \mathcal{K}$ contains an option x such that $\mathcal{P}(x) = \mathcal{P}(x')$ for another option $x' \in X$ (a *clone* of x), there is a context $K' \in \mathcal{K}$ such that $K' = K \cup \{x'\}$. This is a weak condition.[8]

Theorem 2 *Given a set of contexts $\mathcal{K}$ that is closed under cloning, a choice function C is reason-based rationalizable with context-invariant and context-unregarding motivation if and only if it satisfies Axioms 1*, 2*, and 3*.*

In fact, Axiom 3* alone is equivalent to rationalizability of choice by a binary relation over options, as is well-known in the classical case where contexts are feasible sets (Richter 1971 and Bossert and Suzumura 2010).

Remark 1 *A choice function C satisfies Axiom 3* if and only if it is rationalizable by a preference relation, i.e., there is a binary relation $\succsim$ on X such that for all contexts $K \in \mathcal{K}$,*

$$C(K) = \{x \in K : x \succsim y \text{ for all } y \in K\}.$$

This, however, is not a *reason-based* rationalization, and to obtain such a rationalization, our additional axioms, 1* and 2*, are needed, as Theorem 2 shows.

3.3 Reason-based rationalizability with either context-unregarding or context-invariant motivation

We finally turn to reason-based rationalizability with one but not both kinds of context-dependence. We begin with the case in which the agent's motivation can be context-variant, but not context-regarding. The axioms characterizing this case lie logically between (i) Axioms 1*, 2*, and 3*, which characterize reason-based rationalizability without any context-dependence (Theorem 2), and (ii) Axioms 1, 2, and 3, which characterize reason-based rationalizability *simpliciter* (Theorem 1). Specifically, they are Axioms 1* and 3 and a new axiom that weakens Axiom 2* in the presence of 1*. We omit the details here, since the new axiom has a complex form.

[8] It holds vacuously if no two distinct options in X have the same properties, i.e., for any $x, x' \in X$, $x \neq x'$ implies $\mathcal{P}(x) \neq \mathcal{P}(x')$. The condition is also natural because if an option x' is property-wise indistinguishable from a currently feasible option x, one would expect that x' can become feasible too. Presumably, if x, but not x', can be feasible (together with some other options), this difference stems from x and x' having different properties. We could further weaken or modify the condition, e.g., by replacing "$K' = K \cup \{x'\}$" with "$K' = (K \backslash \{x : \mathcal{P}(x) = \mathcal{P}(x')\}) \cup \{x'\}$", so that x' is not added but substituted for the existing feasible options that are property-wise indistinguishable from it.

Now consider the case of context-invariant but possibly context-regarding motivation, which subsumes sophisticated rational behaviour, as illustrated by Polite Pauline. Surprisingly, the conditions characterizing this case are the same as those characterizing reason-based rationalizability without any restrictions. Thus, any choice behaviour that is reason-based rationalizable also has a rationalization with context-invariant motivation. Although this suggests that the restriction to context-invariance has no choice-behaviourial implications, we show in Section 5 that this impression is misleading. The restriction to context-invariance can affect the prediction of choices in novel contexts.

Before stating the present result formally, let us give an illustration. As we have seen, Chocoholic Coco can be rationalized by a reason-based model with context-variant motivation. This captures our informal description of Coco's behaviour. However, a less intuitive rationalization is also possible. It ascribes context-*in*variant motivation to Coco, at the expense of making this motivation context-regarding. This alternative model $(M, \geq)$ is the following:

- M assigns to each context the same set of motivationally salient properties $M = \{\text{big, medium, small, chocolate-offering}\}$, instead of letting motivationally salient properties vary with the presence or absence of chocolate;
- $\geq$ places any property bundles that do not contain the property "chocolate-offering" in the same indifference class (e.g., $\{\text{big}\} \equiv \{\text{small}\}$), and ranks property bundles by size when they contain one of the size properties together with the property "chocolate-offering" (e.g., $\{\text{big, chocolate-offering}\} > \{\text{medium, chocolate-offering}\} > \{\text{small, chocolate-offering}\}$).

Generally, two reason-based models $\mathcal{M}$ and $\mathcal{M}'$ are *behaviourally equivalent* if they induce the same (possibly improper) choice function, i.e., if $C^{\mathcal{M}} = C^{\mathcal{M}'}$.

Proposition 1 *Every reason-based model is behaviourally equivalent to one with context-invariant motivation.*

Corollary 1 *A choice function C has a reason-based rationalization with context-invariant motivation if and only if it has a reason-based rationalization simpliciter.*

The possibility of re-modelling any reason-based rationalization in a context-invariant way disappears once we impose further requirements on $\mathcal{M}$, such as the requirement that motivation be context-unregarding or that it be "revealed", as discussed in Section 4.[9] As a consequence of Proposition 1, Theorem 1 can be re-stated as a characterization of context-invariant reason-based choice:

[9]Even when this re-modelling is possible, it may sacrifice parsimony and psychological adequacy, as

Theorem 3 *A choice function C is reason-based rationalizable with context-invariant motivation if and only if it satisfies Axioms 1 and 3 (and by implication 2).*

3.4 Criteria for selecting a rationalization in cases of non-uniqueness

How can we select a reason-based model $(M, \geq)$ in cases of non-uniqueness?[10] This question matters because different models attribute to the agent different cognitive processes, which may differ in psychological adequacy and lead to different predictions for the agent's future choices, as discussed in Section 5. There are at least three kinds of criteria for selecting a model.

Revelation criteria: These require that, as far as possible:

(i) the motivational salience function M deem only those properties motivationally salient that make an observable difference to the agent's choice behaviour, and

(ii) the fundamental preference relation $\geq$ over property bundles be systematically derived from the agent's choice behaviour.

The goal is to minimize behaviourally ungrounded ascriptions of motivation and fundamental preference. This is the topic of Section 4.

Non-choice data: Verbal reports or neurophysiological data, such as responses to property-related stimuli, may help us test hypotheses about

(i) which properties are motivationally salient for the agent in context K and thus belong to $M(K)$,

(ii) which context properties causally affect motivational salience, so that $M(K)$ may vary as contexts K vary in those properties, and

(iii) which property bundles the agent fundamentally prefers to which others.

---

evident from the proof of Proposition 1. Here, *every* property that was motivationally salient in *some* context in the original, context-variant model $(M, \geq)$ and every context property (at least every context property on which $M(K)$ may depend) becomes motivationally salient in the new, context-invariant model $(M^*, \geq^*)$, with M^* constant. Formally, $(\cup_{K \in \mathcal{K}} M(K)) \cup \mathcal{P}_{\text{context}} \subseteq M^*$.

[10] Non-uniqueness in the rationalization of choice behaviour is familiar from classical choice theory, where the same choice function can often be rationalized by more than one binary relation over the options. The relation becomes unique if the domain of the choice function (i.e., the set of contexts in which choice is observed) is "rich", i.e., contains all sets of one or two options.

One might hypothesize that people have better conscious access to how they conceptualize the options in a given context K and therefore to the properties in $M(K)$ than to the context properties that affect what $M(K)$ is (i.e., those properties which, in an empirical study, might be significant explanatory variables for M). Some changes in $M(K)$ might be due to subconscious influences, as in framing or nudging effects. If so, verbal reports may be more relevant to questions (i) and (iii) than to question (ii).

Parsimony criteria: We may try to select a *parsimonious* model $(M, \geq)$, where

(i) the sets $M(K)$ of motivationally salient properties generated by M are (a) as small as possible and (b) as unchanging as possible across different K, and

(ii) the relation $\geq$ is as sparse as possible (e.g., defined over the fewest possible property bundles).

There may be a trade-off between different dimensions of parsimony. If the sets $M(K)$ contain only few properties, they may not be stable across different K, and vice versa. As the proof of Proposition 1 shows, we can always achieve context-invariance by defining M constantly as the entire set $\mathcal{P}$ and the fundamental preference relation $\geq$ as the revealed preference relation $\succsim^C$ over property bundles. This makes the sets $M(K)$ unchanging but very large, and hence perhaps psychologically implausible. Conversely, making each $M(K)$ small might require variation across contexts.

4 The revealed reason-based model

A familiar concept from classical choice theory is the revealed preference relation over options, which can be inferred from the agent's choice behaviour. Analogously, we now introduce the revealed reason-based model, which can be inferred from the observed choice function. It is constructed by

- counting a property as motivationally salient in a given context if and only if it makes a behavioural difference (in a sense defined below), and
- counting a property bundle S as fundamentally preferred to another bundle T if and only if the agent chooses an option x over another option y, where x and y are revealed to be conceptualized as S and T (in a sense defined below).

We first define the revealed reason-based model and then characterize the class of choice functions that are rationalizable by such a model.

4.1 Revealed motivationally salient properties

Our strategy for determining whether a property P is motivationally salient for an agent in context K is to ask whether its presence or absence in an option makes a difference to his choice in contexts "like" K, i.e., contexts K' with the same context properties as K (where $\mathcal{P}(K')=\mathcal{P}(K)$). Choices in contexts with different context properties are irrelevant, since they could stem from different motivationally salient properties. The choice of moisturizer over sunscreen in a cloudy context is no evidence for whether "protecting against UV radiation" is motivationally salient in a context with sunshine.

To formalize these ideas, we begin with some preliminary terminology. Two property bundles *agree* on a property $P \in \mathcal{P}$ if both or neither contain P; otherwise, they *differ* in P. A property bundle S is *weakly between* two property bundles T and T' if S agrees with each of T and T' on every property on which they agree. If, in addition, S is distinct from each of T and T', then S is *strictly between* T and T'. For instance, $\{P, Q\}$ is strictly between $\{P\}$ and $\{Q\}$, as is $\varnothing$. Two property bundles are *revealed comparable* if one of them is chosen in some context K while the other is feasible. Two such bundles *differ minimally* if there is no property bundle that is strictly between them and revealed comparable to at least one of them.

One might think that a property P is motivationally salient in context K if and only if there is a context K' with the same context properties as K in which the agent reveals a strict preference between two property bundles that differ in P. But this criterion is inadequate, because the two bundles may also differ in other properties. The agent may choose the larger of two T-shirts, not because it is larger, but because it is blue. So, before we can infer that P is motivationally salient, we must verify that the two property bundles differ *minimally*. This suggests the following criterion.

Criterion 1 *Property P is revealed motivationally salient in context K if there exist property bundles S and S' such that*

(rev1) S and S' differ in P,

(rev2) S is revealed strictly preferred to S' or vice versa, where the contexts in which S and S' are feasible have the same context properties as K (i.e., $S \cap \mathcal{P}_{context} = S' \cap \mathcal{P}_{context} = \mathcal{P}(K)$), and

(rev3) S and S' differ minimally.

However, this criterion excludes some natural cases. Suppose, again, the options are T-shirts, and P is the property of largeness. If every context offers either only large T-shirts or only small ones, P cannot satisfy Criterion 1, since no revealed comparable sets

S and S' ever satisfy (rev2). But suppose that whenever only large T-shirts are available the agent chooses the darkest, and whenever only small T-shirts are available he chooses the lightest. If there are no context properties in $\mathcal{P}$ by which we could distinguish those contexts further and to which we could attribute the behavioural difference, it is natural to conclude that property P is motivationally salient. This is because the agent's choice between two property bundles containing the property "large" (a large dark T-shirt and a large light one) is reversed when we remove that property from them (so that we are now comparing a small dark T-shirt and a small light one). These considerations suggest the following more general criterion, by which we define *revealed motivational salience*.

Criterion 2 *Property P is revealed motivationally salient in context K if there exist two pairs of property bundles (S,T) and (S',T') such that*

(REV1) the two pairs differ in P, i.e., S and S' differ in P, or T and T' differ in P,

(REV2) S is revealed preferred to T while T' is revealed preferred to S' or vice versa (with at least one preference strict), where the contexts in which S and T, or S' and T', are feasible have the same context properties as K (i.e., $S \cap \mathcal{P}_{context} = S' \cap \mathcal{P}_{context} = T \cap \mathcal{P}_{context} = T' \cap \mathcal{P}_{context} = \mathcal{P}(K)$), and

(REV3) the pair (S,T) differs minimally from the pair (S',T'), i.e., there is no other pair (S'',T'') (with S'' revealed comparable to T'') such that S'' is weakly between S and S' and T'' is weakly between T and T'.

In our example, S and T could be the property bundles instantiated by the large dark T-shirt and the large light T-shirt, and S' and T' the bundles instantiated by the small dark T-shirt and the small light T-shirt, respectively.

Proposition 2 *Criterion 2 generalizes Criterion 1, i.e., for any context $K \in \mathcal{K}$, any property $P \in \mathcal{P}$ that satisfies (rev1)-(rev3) (for some $S, S' \subseteq \mathcal{P}$) also satisfies (REV1)-(REV3) (for some $S, S', T, T' \subseteq \mathcal{P}$).*

Our definition of revealed motivational salience has the following natural implication:

Lemma 2 *(informal statement) The revealed preference between any two revealed comparable property bundles S and T (i.e., whether $S \succsim^C T$) depends only on*

- *the context properties within S and T (these determine the contexts K in which S and T are feasible), and*
- *the properties within S and T that are revealed motivationally salient in such contexts K.*

4.2 The revealed reason-based model

We can now complete our definition of the revealed reason-based model. The *revealed motivational salience function* is the function M^C (from $\mathcal{K}$ into $2^{\mathcal{P}}$) satisfying:

for each context K, $M^C(K) = \{P \in \mathcal{P} : P \text{ is revealed motivationally salient in } K\}$.

To illustrate, the revealed motivational salience functions of the four agents in our example above – Bonnie, Pauline, Coco, and William – are precisely the motivational salience functions that we used to rationalize their choices.[11]

Given the function M^C, any option x is *revealed conceptualized* in context K as

$$x_K^C = \mathcal{P}(x, K) \cap M^C(K).$$

We define a property bundle S to be *revealed weakly fundamentally preferred* to another property bundle T, denoted $S \geq^C T$, if, in some context $K \in \mathcal{K}$, there are feasible options x and y, revealed conceptualized as $x_K^C = S$ and $y_K^C = T$, such that $x \in C(K)$.

The model $(M^C, \geq^C)$ is called the *revealed reason-based model*. It can be checked that the reason-based models that we used to rationalize the four agents in our example are indeed the revealed models. By considering the revealed model for an agent, we can behaviourally determine which, if any, of the two kinds of context-dependence are present in the agent's motivation. In our example, Coco and William have revealed context-variant motivation, while Bonnie and Pauline do not; and Pauline and William have revealed context-regarding motivation, while Bonnie and Coco do not.

4.3 Rationalizability by the revealed model

Is every reason-based rationalizable choice function also rationalizable by the revealed model? Recall that reason-based rationalizability *simpliciter* requires Axioms 1 and 3 (which, in turn, imply Axiom 2). For rationalizability by the revealed model, we must strengthen these axioms by adding the following variant of Axiom 2.

[11] For instance, for Bonnie, to show that big$\in M^C(K)$ for any K without chocolate-covered pears, check (rev1)-(rev3) for S={big} and S'={medium}; to show that big$\in M^C(K)$ for any K with chocolate-covered pears, check (rev1)-(rev3) for S={big,chocolate-offering} and S'={medium,chocolate-offering}. For Pauline, to show that polite$\in M^C(K)$ for any K without chocolate-covered pears, check (rev1)-(rev3) for S={big,polite} and S'={big}; to show this for any K with chocolate-covered pears, check (rev1)-(rev3) for S={big,polite,chocolate-offering} and S'={big,chocolate-offering}. Strictly speaking, the sets $M^C(K)$ take this form if X is sufficiently rich, i.e., contains fruits instantiating relevant property bundles.

Axiom 2** For all contexts $K, K' \in \mathcal{K}$, if $\{x_K^C : x \in K\} = \{x_{K'}^C : x \in K'\}$, then $\{x_K^C : x \in C(K)\} = \{x_{K'}^C : x \in C(K')\}$.

Our theorem requires a technical condition. Call the set $\mathcal{K}$ of contexts *rich* if, whenever two property bundles S and T are simultaneously feasible in some context in $\mathcal{K}$, then $\mathcal{K}$ contains a context in which *only* S and T are feasible.

Theorem 4 *Given a rich set of contexts $\mathcal{K}$, a choice function C is rationalizable by the revealed reason-based model $(M^C, \geq^C)$ if and only if it satisfies Axioms 1, 2**, and 3.*[12]

Surprisingly, Theorem 4 does not explicitly require the following variant of Axiom 1.

Axiom 1** For all contexts $K \in \mathcal{K}$ and all options $x, y \in K$, if $x_K^C = y_K^C$, then $x \in C(K) \Leftrightarrow y \in C(K)$.

Lemma 3 *Axioms 1 and 1** are equivalent.*

To see that rationalizability by the revealed model is more demanding than reason-based rationalizability *simpliciter*, we give an example.[13] Suppose the options are electoral candidates, and the contexts are elections. Let $\mathcal{K} = \{K_1, K_2\}$, and consider an agent who in context K_1 votes for any candidate with the (option) property "experienced" (say, over 20 years of political experience) and in context K_2 votes for any candidate with the (option) property "young" (say, aged below 50), where candidates of both kinds are available in both contexts. This choice behaviour can be rationalized by a reason-based model $(M, \geq)$ in which $M(K_1) = \{\text{experienced}\}$ and $M(K_2) = \{\text{young}\}$, and $\geq$ satisfies

$$\{\text{experienced}\} > \varnothing \text{ and } \{\text{young}\} > \varnothing.$$

What is the revealed model? Suppose there is a perfect anti-correlation between the properties "experienced" and "young": a candidate in X is experienced if and only if he or she is not young. We then have no choice-behavioural basis for determining whether

[12] We may further ask whether a given choice function C is rationalizable by a model $(M^C, \geq)$ in which M^C is the revealed motivational salience function but $\geq$ is unrestricted. In the Appendix, we prove that, given a rich $\mathcal{K}$, a choice function C is rationalizable by some such model if and only if it satisfies Axioms 1, 2**, and 3. Further, this model $(M^C, \geq)$ will be essentially identical to the revealed model $(M^C, \geq^C)$. Two models $(M, \geq)$ and $(M', \geq')$ are *essentially identical* if (i) $M = M'$, and (ii) the fundamental preference relations $\geq$ and $\geq'$ coincide wherever they are choice-behaviourally relevant (i.e., $S \geq T \Leftrightarrow S \geq' T$ for all property bundles S and T such that there are options x and y in some context K that are conceptualized as $\mathcal{P}(x, K) \cap M(K) = S$ and $\mathcal{P}(y, K) \cap M(K) = T$, respectively).

[13] The point that choice may be rationalizable, but not by the revealed model, arises also in classical choice theory: if we seek to rationalize choice by a complete and transitive preference relation, there may exist such a rationalization although the revealed preference relation is neither complete nor transitive.

"experienced" or "young" or both are motivationally salient for our voter in any context: he might have voted for an experienced candidate in context K_1, not because he cares about (and likes) experience in politicians, but because he cares about (and dislikes) youth. Formally, both properties are *revealed* motivationally salient in contexts K_1 and K_2. We have $M^C(K_1) = M^C(K_2) = \{\text{experienced, young}\}$.[14]

It is impossible to rationalize the present choice behaviour by the revealed reason-based model $(M^C, \geq^C)$ or any other model of the form $(M^C, \geq)$. According to M^C, our voter always conceptualizes every candidate either as {experienced} or as {young}, where his choice in context K_1 can only be rationalized if {experienced} > {young}, while his choice in context K_2 can only be rationalized if {young} > {experienced}. It is easy to check that the voter's choice behaviour violates Axiom 2**.[15]

5 Predicting choices in novel contexts

Standard choice theory is largely silent on how to predict choices in novel, previously unobserved contexts. In almost every empirical science, we make predictions about future (or otherwise unobserved) events, based on past observations. Astronomers predict future solar eclipses or movements of comets based on past trajectories of the relevant celestial bodies; epidemiologists predict future epidemics based on past epidemiological data; and econometricians use past data of the economy to predict its future. Choice theory is an exception in that predictions and observations are usually taken to be the same thing: the choice function is the observed *and* predicted object at once.

Genuine predictions would have to be about choice contexts outside the observed domain $\mathcal{K}$, perhaps with feasible options outside the set X. If we rationalize choices simply by a preference relation on X, we have no systematic way of extending this relation to new options. So, we can make only two rather trivial kinds of predictions:

- Any choice function on a set $\mathcal{K}$ of contexts can predict choices when contexts in $\mathcal{K}$ recur in the future. But here the preference relation does no work, since even a not-yet-rationalized choice function allows us to make the same predictions.

- A preference relation on X might be used to predict choices in contexts outside $\mathcal{K}$ that involve only "old" options from X. In such "slightly novel" contexts, we

[14] We assume that $\mathcal{P}$ contains only the option properties "experienced" and "young" and some context properties to which the change in motivation from K_1 to K_2 can be attributed.

[15] Although $\{x^C_{K_1}{:}x{\in}K_1\}{=}\{x^C_{K_2}{:}x{\in}K_2\}{=}\{\{\text{experienced}\},\{\text{young}\}\}$, we have $\{x^C_{K_1}{:}x{\in}C(K_1)\} \neq \{x^C_{K_2}{:}x{\in}C(K_2)\}$, since $\{x^C_{K_1}{:}x{\in}C(K_1)\}{=}\{\{\text{experienced}\}\}$ and $\{x^C_{K_2}{:}x{\in}C(K_2)\}{=}\{\{\text{young}\}\}$.

would predict that the agent will maximize the same preference relation over the feasible options.

To introduce a reason-based approach to predictions in genuinely new contexts, we present a formal framework and explore predictions of more and less conservative kinds.

5.1 A framework for predictions

We take the options in X, the contexts in $\mathcal{K}$, and the choice function C to refer to *previously observed* choices, and introduce some further primitives:

- An extended set $X^+ \supseteq X$ of options, containing additional options the agent might encounter.
- An extended set $\mathcal{K}^+ \supseteq \mathcal{K}$ of contexts, containing additional choice contexts the agent might encounter. Every "new" context K (in $\mathcal{K}^+ \backslash \mathcal{K}$), like every "old" one (in $\mathcal{K}$), induces a non-empty set $[K]$ of feasible options. Again, we write K for $[K]$ when there is no ambiguity. While in "old" contexts only "old" options (in X) are feasible, in "new" contexts "new" options (in $X^+ \backslash X$) can be feasible.
- The agent's extended choice function C^+ on $\mathcal{K}^+$. This is an extension of the observed choice function C (i.e., the restriction of C^+ to $\mathcal{K}$ coincides with C) and is interpreted as the "true" choice function, capturing the choices the agent would make when confronted with the contexts in $\mathcal{K}^+$.

Having observed the agent's choices in the domain $\mathcal{K}$, we wish to predict his choices in $\mathcal{K}^+$. The goal is to predict as much of the "true" choice function C^+ as possible. A *choice predictor* is a choice function π on some domain $\mathcal{D} \subseteq \mathcal{K}^+$, where typically $\mathcal{K} \subseteq \mathcal{D} \subseteq \mathcal{K}^+$. For each K in $\mathcal{D}$, $\pi(K)$ is the predicted choice in context K. The predictor is *accurate* if it predicts the agent's choice correctly in all contexts in $\mathcal{D}$, i.e., if $\pi(K) = C^+(K)$ for all K in $\mathcal{D}$. As noted above, a preference relation on X would only allow us to define predictors for "old" contexts $K \in \mathcal{K}$ or for "new" contexts $K \notin \mathcal{K}$ containing only "old" options from X. Reason-based rationalizations allow us to go further.

We now assume that the properties in $\mathcal{P}$ are defined over the extended set of option-context pairs $X^+ \times \mathcal{K}^+$ (not just over the pairs in $X \times \mathcal{K}$). For any domain of contexts $\mathcal{D} \subseteq \mathcal{K}^+$, a *reason-based model for domain* $\mathcal{D}$ is defined as before and again denoted $(M, \geq)$, but ranges over $\mathcal{D}$ instead of $\mathcal{K}$. In particular, M is a function from $\mathcal{D}$ into $2^{\mathcal{P}}$. Our strategy for defining a choice predictor is the following:

- Take a reason-based model $\mathcal{M} = (M, \geq)$ for the original domain $\mathcal{K}$ as given.

- Extend this to a model $\mathcal{M}' = (M', \geq)$ for some domain $\mathcal{D}$ with $\mathcal{K} \subseteq \mathcal{D} \subseteq \mathcal{K}^+$.
- Define a choice predictor on $\mathcal{D}$ as the choice function $\pi := C^{\mathcal{M}'}$ induced by the extended model.

By an *extension* of the model $\mathcal{M} = (M, \geq)$ to the domain $\mathcal{D} \supseteq \mathcal{K}$ we mean a reason-based model $\mathcal{M}' = (M', \geq)$ for domain $\mathcal{D}$ whose restriction to $\mathcal{K}$ is $\mathcal{M}$, i.e., (i) the restriction of the function M' to the subdomain $\mathcal{K}$ is M, and (ii) $\mathcal{M}$ and $\mathcal{M}'$ use the same fundamental preference relation $\geq$.

5.2 Cautious, semi-courageous, and courageous prediction

We now define three reason-based choice predictors. Each is based on a reason-based model $\mathcal{M} = (M, \geq)$ by which we have rationalized the agent's observed choice. (This could be, for example, the revealed model $(M^C, \geq^C)$ discussed in Section 4.)

Cautious prediction: The *cautious predictor* (based on $\mathcal{M}$) is the choice function $\pi := C^{\mathcal{M}'}$ induced by the extended model $\mathcal{M}' = (M', \geq)$ whose domain $\mathcal{D}$ consists of every context $K \in \mathcal{K}^+$ such that K offers the same feasible property bundles as some observed context $L \in \mathcal{K}$:

$$\{\mathcal{P}(x, K) : x \in K\} = \{\mathcal{P}(x, L) : x \in L\}. \tag{1}$$

Note that (1) implies $\mathcal{P}(K) = \mathcal{P}(L)$, so that $M(K)$ must equal $M(L)$. By implication, the extension $\mathcal{M}'$ of $\mathcal{M}$ is uniquely defined.

The cautious predictor makes predictions only for choice contexts that offer the same feasible property bundles as some observed context. This ignores the fact that reason-based choices depend only on motivationally salient properties. The cautious predictor cannot predict, for example, Bonnie's choices from a "new" fruit basket (in $\mathcal{K}^+ \backslash \mathcal{K}$) that is identical to some "old" basket (in $\mathcal{K}$) in terms of the sizes of available fruit but not in terms of other, non-salient properties. We now introduce a predictor based not on entire property bundles but only on bundles of motivationally salient properties.

Semi-courageous prediction: The *semi-courageous predictor* (based on $\mathcal{M}$) is the choice function $\pi := C^{\mathcal{M}'}$ induced by the extended model $\mathcal{M}' = (M', \geq)$ whose domain $\mathcal{D}$ consists of every context $K \in \mathcal{K}^+$ such that

(i) K has the same context properties as some observed context, i.e., $\mathcal{P}(K) = \mathcal{P}(L)$ for some L in $\mathcal{K}$ (so that $M(K) = M(L)$), and

(ii) the set of *options as conceptualized in* K (feasible bundles of *motivationally salient* properties) is the same as that in some observed context, i.e., $\{x_K : x \in K\} = \{x_L : x \in L'\}$ for some L' in $\mathcal{K}$.

Note that L and L' in clauses (i) and (ii) can be distinct. Although the semi-courageous predictor can predict choices in contexts offering new feasible property bundles, it is still somewhat restrictive. Clause (i) is often unnecessarily demanding. Its role is to tell us how we must define $M(K)$, namely as $M(L)$. Sometimes, however, we can infer how to define $M(K)$ without clause (i). Consider, for example, an agent with context-invariant motivation (according to $\mathcal{M}$). If we are willing to assume that his motivation remains context-invariant in new contexts, we can define $M(K)$ as unchanged outside $\mathcal{K}$. This suggests the following, more general predictor.

Courageous prediction: We begin with a preliminary definition. In a reason-based model $\mathcal{M}' = (M', \geq)$ for some domain $\mathcal{D}$, we call a context property P *causally relevant* if its presence or absence in a context can make a difference to the agent's set of motivationally salient properties in it, i.e., if there are contexts $K, K' \in \mathcal{D}$ such that

(cau1) K has property P while K' does not (or vice versa),

(cau2) K and K' induce different sets of motivationally salient properties, i.e., $M'(K) \neq M'(K')$,

(cau3) K and K' differ minimally, i.e., there is no context $K'' \in \mathcal{D}$ whose set of context properties $\mathcal{P}(K'')$ is strictly between the sets $\mathcal{P}(K)$ and $\mathcal{P}(K')$.[16]

Let $CAU^{\mathcal{M}'}$ denote the set of causally relevant context properties in model $\mathcal{M}'$.[17] Two things are worth noting. First, in the important special case of context-invariant motivation, *no* context property is causally relevant. Second, the causally relevant context properties fully determine the agent's set of motivationally salient properties. Formally:

Proposition 3 *Let $\mathcal{M}' = (M', \geq)$ be any reason-based model (for some domain $\mathcal{D}$ of contexts). Then:*

(a) $\mathcal{M}'$ has context-invariant motivation if and only if $CAU^{\mathcal{M}'} = \varnothing$.

(b) For all K, K' in $\mathcal{K}$, if $\mathcal{P}(K) \cap CAU^{\mathcal{M}'} = \mathcal{P}(K') \cap CAU^{\mathcal{M}'}$ then $M'(K) = M'(K')$.

[16] This clause excludes the possibility that K and K' differ in context properties unrelated to P to which the difference in motivation between K and K' could be causally attributed.

[17] If $\mathcal{M}'$ is a model with revealed motivation (i.e., $\mathcal{M}' = (M^C, \geq)$), causal relevance is fully determined by the observed choice function C, so that we may speak of *revealed causal relevance* and write CAU^C.

The *courageous predictor* (based on $\mathcal{M}$) is the choice function $\pi := C^{\mathcal{M}'}$ induced by the extended model $\mathcal{M}' = (M', \geq)$ whose domain $\mathcal{D}$ consists of every context $K \in \mathcal{K}^+$ such that

(i*) K has the same causally relevant properties as some observed context, i.e., $\mathcal{P}(K) \cap CAU^{\mathcal{M}} = \mathcal{P}(L) \cap CAU^{\mathcal{M}}$ for some L in $\mathcal{K}$; we then define $M(K)$ as $M(L)$;[18] and

(ii) the set of *options as conceptualized in* K is the same as that in some observed context, i.e., $\{x_K : x \in K\} = \{x_L : x \in L'\}$ for some L' in $\mathcal{K}$.

Our three predictors are increasingly general:

Remark 2 *Given a reason-based rationalization $\mathcal{M}$ of the observed choice function C,*

(a) the cautious predictor extends the observed choice function C;

(b) the semi-courageous predictor extends the cautious predictor; and

(c) the courageous predictor extends the semi-courageous predictor.[19]

5.3 When is each choice predictor accurate?

When does each predictor coincide with the true choice function C^+ on its domain? It turns out that the accuracy of each predictor depends on whether certain observed patterns in the agent's choices are *robust*, i.e., continue to hold in contexts outside $\mathcal{K}$.

Theorem 5 *Given a reason-based rationalization $\mathcal{M}$ of the observed choice function C,*

(a) the cautious predictor is accurate if the extended choice function C^+ is rationalizable by some reason-based model;

(b) the semi-courageous predictor is accurate if the extended choice function C^+ is rationalizable by some extension of $\mathcal{M}$; and

(c) the courageous predictor is accurate if the extended choice function C^+ is rationalizable by some extension of $\mathcal{M}$ with the same causally relevant context properties.

[18] By Proposition 3, the definition of $M(K)$ does not depend on the choice of L.

[19] The three predictors could be extended further in analogy to the second route we mentioned for predictions based on preference relations alone: we could drop the requirement that any context K in $\mathcal{D}$ must offer the same feasible property bundles (in the cautious case) or options-as-conceptualized (in the other cases) as some context in $\mathcal{K}$. The maximal generalization would replace clause (ii) in the definition of the courageous predictor with the requirement that $\{x_K : x \in K\}$ has a $\geq$-greatest element.

Informally, part (a) shows that cautious predictions are accurate if the agent's choices are robustly reason-based, i.e., reason-based not just in the observed domain $\mathcal{K}$ but also in the extended domain $\mathcal{K}^+$. This seems plausible for agents with some stability in their choice dispositions. Part (b) shows that semi-courageous predictions are accurate if the existing model $\mathcal{M}$ rationalizes choice robustly: it not only explains the agent's *observed* choices, but can be extended to explain all new choices too. This requires that our reason-based model for the observed domain $\mathcal{K}$ is a portion of a reason-based model for the extended domain $\mathcal{K}^+$. Part (c) shows that courageous predictions are accurate if the model $\mathcal{M}$ rationalizes choice robustly in a stronger sense: its extension to new contexts requires no additional causally relevant context properties. So, our reason-based model for $\mathcal{K}$ must be a portion of a reason-based model for the extended domain $\mathcal{K}^+$ that already identifies all causally relevant context properties. Whether these robustness assumptions are justified depends, in part, on how rich the domain $\mathcal{K}$ of observed contexts is relative to the target domain $\mathcal{K}^+$. Let us explain this in relation to our three-part theorem:

(a) If the observed domain $\mathcal{K}$ is small, then reason-based rationalizability in $\mathcal{K}$ is only limited evidence for reason-based rationalizability in the larger domain $\mathcal{K}^+$. In the limit, if $\mathcal{K}$ contains only contexts with singleton feasible sets, the agent's choices are trivially reason-based rationalizable in $\mathcal{K}$, and we have no evidence for reason-based rationalizability in $\mathcal{K}^+$. By contrast, if $\mathcal{K}$ contains a large and representative mix of contexts – e.g., a sizeable "random sample" of contexts from $\mathcal{K}^+$ – then reason-basedness in $\mathcal{K}$ may be good evidence for reason-basedness in $\mathcal{K}^+$.

(b) Even if the agent's choices are robustly reason-based, our reason-based model for $\mathcal{K}$ need not be a portion of a model for $\mathcal{K}^+$. The set $M(K)$ specified for some observed context K may leave out some property that is needed to explain the agent's choice in some new context K' with $\mathcal{P}(K') = \mathcal{P}(K)$. If so, a reason-based model for $\mathcal{K}^+$ could not be an extension of our model for $\mathcal{K}$, since it would have to specify the same $M(K') = M(K)$ for *all* contexts K' with $\mathcal{P}(K') = \mathcal{P}(K)$. The larger and more representative $\mathcal{K}$ is, the less likely this problem is to occur.

(c) Similar remarks apply to the question of whether our model for $\mathcal{K}$, even if it is extendible to a model for $\mathcal{K}^+$, is likely to identify all context properties that are causally relevant in $\mathcal{K}^+$. For example, if $\mathcal{K}$ contains no choice contexts offering luxury goods, then our model for $\mathcal{K}$ cannot identify the difference that "offering luxury goods" might make to the agent's motivation in contexts with that property. A large and representative domain $\mathcal{K}$ reduces the risk of not identifying some context properties that are causally relevant in the target domain $\mathcal{K}^+$.

6 Concluding remarks

We have argued that reason-based rationalizations can explain a variety of non-classical choice behaviours in a unified manner and clarify the difference between "bounded" and "sophisticated" deviations from classical rationality. Furthermore, unlike classical choice-theoretic rationalizations in terms of preference relations over options, they allow us to predict an agent's choices in genuinely novel contexts, where no observations have been made. Crucially, different rationalizations of the same choice behaviour are not generally equivalent, since some are typically more likely than others to extend robustly to new choice contexts and thus to lead to accurate predictions of future choices.

Such robustness is related to psychological adequacy. A psychologically ungrounded explanation of an agent's observed choices is more likely to "fail" in novel contexts, because it matches the observations by coincidence rather than for systematic reasons that continue to apply in novel contexts. Psychological adequacy thus matters for the sake of predictive accuracy, regardless of whether it matters for its own sake.

7 References

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

Bhattacharyya, A., P. K. Pattanaik, and Y. Xu (2011) "Choice, Internal Consistency and Rationality", *Economics and Philosophy* 27(2): 123-149.

Bossert, W., and K. Suzumura (2009) "External Norms and Rationality of Choice", *Economics and Philosophy* 25: 139-152.

Bossert, W., and K. Suzumura (2010) *Consistency, Choice and Rationality.* Cambridge/MA (Harvard University Press).

Cherepanov, V., T. Feddersen, and A. Sandroni (2013) "Rationalization", *Theoretical Economics* 8(3): 775-800.

Dietrich, F. (2012) "Modelling change in individual characteristics: an axiomatic framework", *Games and Economic Behavior* 76: 471-494.

Dietrich, F., and C. List (2013a) "A reason-based theory of rational choice", *Nous* 47(1): 104-134.

Dietrich, F., and C. List (2013b) "Where do preferences come from?", *International Journal of Game Theory* 42(3): 613-637.

Dietrich, F., and C. List (2012) "Mentalism versus behaviourism in economics: a philosophy-of-science perspective", ms., London School of Economics.

Kalai, G., Rubinstein, A., and R. Spiegler (2002) "Rationalizing choice functions by multiple rationales", *Econometrica* 70: 2481-2488.

Lancaster, K. J. (1966) "A new approach to consumer theory", *Journal of Political Economy* 74(2): 132-157.

Liu, F. (2010) "Von Wright's 'The Logic of Preference' revisited", *Synthese* 175(1): 69-88.

Mandler, M., P. Manzini, and M. Mariotti (2012) "A million answers to twenty questions: Choosing by checklist", *Journal of Economic Theory* 147: 71-92.

Manzini, P., and M. Mariotti (2007) "Sequentially rationalizable choice", *American Economic Review* 97(5): 1824-1839.

Manzini, P., and M. Mariotti (2012) "Moody choice", ms., University of St Andrews.

Masatlioglu, Y., D. Nakajima, and E. Y. Ozbay (2012) "Revealed Attention", *American Economic Review* 102(5): 2183-2205.

Osherson, D., and S. Weinstein (2012) "Preferences based on reasons", *The Review of Symbolic Logic* 5(1): 122-147.

Pettit, P. (1991) "Decision Theory and Folk Psychology", in M. Bacharach and S. Hurley (eds.), *Foundations of Decision Theory.* Oxford (Blackwell): 147-175.

Richter, M. (1971) "Rational Choice", in J. Chipman et al. (eds.), *Preferences, Utility and Demand.* New York (Harcourt Brace Jovanovich): 29-58.

Rubinstein, A. (2006) *Lecture Notes in Microeconomic Theory: The Economic Agent.* Princeton (Princeton University Press).

Salant, Y., and A. Rubinstein (2008) "(A, f): Choice with Frames", *Review of Economic Studies* 75: 1287-1296.

Samuelson, P. (1948) "Consumption theory in terms of revealed preferences", *Economica* 15(60): 243-253.

Suzumura, K., and Y. Xu (2001) "Characterizations of consequentialism and nonconsequentialism", *Journal of Economic Theory* 101: 423-436.

A Appendix

Before giving the proofs, we briefly discuss related works by Bossert and Suzumura (2009) (for short, B&S) and Bhattacharyya, Pattanaik, and Xu (2011) (for short, B&P&X). B&S assume that, in any given choice context, a feasible option may or may not be compatible with some exogenously given 'norms'. In our example, picking the only available apple would violate a politeness norm. B&S axiomatically characterize those choice functions which are *norm-conditionally rationalizable*: there exists a preference relation over options such that, in any context, the agent chooses the most preferred *norm-compatible* feasible option. One may think of such a rationalization as being 'partially reason-based'. Each norm gives rise to a (context-regarding) property: the property of obeying that norm. Every such property is taken to be desirable and motivationally salient in each context. The agent's choice of a norm-compatible option is then explained by the fact that the option has all those properties (of obeying the norms in question). By contrast, the question of *which* of the norm-compatible options is chosen is not explained in terms of reasons (properties), but in terms of a standard preference relation over primitive options. B&P&X take a different approach. Like us, they model the agent's conceptualization of options, yet not by invoking properties or reasons, but by refining the notion of an option through adding certain 'relevant' information about the context. To describe Polite Pauline in our example, the options (fruits) would have to be refined by including the information of whether or not the context offers another fruit of the same kind. The refinement is carried out by a technical construction.[2] B&P&X show that an agent whose choices among refined options are fully rational may nonetheless 'look' irrational if his choice function is defined over non-refined options. Overall, B&S's and B&P&X's analyses convey several important insights relevant to our paper.

Notation. For property bundles $S, T \subseteq \mathcal{P}$ we write $S \gtrsim\!\!\!\lesssim^C T$ to indicate that S and T are revealed comparable, i.e., that $S \gtrsim^C T$ or $T \gtrsim^C S$. Furthermore, when we need to refer explicitly to the underlying model $\mathcal{M}$, we write $x_K^{\mathcal{M}}$ rather than x_K for option x as conceptualized in context K, and $\gtrsim_K^{\mathcal{M}}$ rather than $\gtrsim_K$ for the induced preference relation in context K. Finally, for brevity, we write M_K rather than $M(K)$ to refer

[2] For B&P&X, a refined option is not simply an option-context pair (x, K) (with $x \in K$), since such an object contains the *full* context information, including any irrelevant information. Rather, B&P&X define refined options as certain *equivalence classes* of such pairs. In the limiting 'classical' case, the context is totally irrelevant, so that any pairs (x, K) and (x, K') count as equivalent; hence, refined options reduce to options in the original sense.

to the set of motivationally salient properties in context K according to motivational salience function M.

Proof of Lemma 1. Assume Axiom 3. As in Axiom 2, consider contexts $K, K' \in \mathcal{K}$ such that (*) $\{\mathcal{P}(y, K) : y \in K\} = \{\mathcal{P}(y', K') : y' \in K'\}$. We only show that $\{\mathcal{P}(x, K) : x \in C(K)\} \subseteq \{\mathcal{P}(x', K') : x' \in C(K')\}$, since the converse inclusion ('$\supseteq$') is analogous. Suppose $x \in C(K)$. The property bundle $\mathcal{P}(x, K)$ is feasible in context K, hence by (*) also in context K'. It is revealed weakly preferred to all feasible property bundles in context K, hence by (*) also to all feasible property bundles in context K'. So, by Axiom 3, it is chosen in context K', i.e., belongs to $\{\mathcal{P}(x', K') : x' \in C(K')\}$. ■

We give no separate proof of Theorem 1, since this result follows from Proposition 1 and Theorem 3, both of which we prove below.

Proof of Theorem 2. Let $\mathcal{K}$ be closed under cloning (an assumption only needed in part 2).

Step 1. Assume C is rationalized by a reason-based model with context-invariant and context-unregarding motivation, $\mathcal{M} = (M, \geq)$, where $M \subseteq \mathcal{P}_{\text{option}}$. We leave the proof of Axioms 1* and 2* to the reader and here prove Axiom 3*. It suffices to show that C is rationalizable in the classical sense by a binary relation on X (see Remark 1). Since $\mathcal{M}$ rationalizes C, the choice set $C(K)$ for a context K consists of the $\succsim_K$-highest option(s) in K, where $\succsim_K^{\mathcal{M}}$ is the preference relation on X induced by the model $\mathcal{M}$ for context K; this relation is defined for all options $x, y \in X$ by

$$x \succsim_K^{\mathcal{M}} y \Leftrightarrow x_K^{\mathcal{M}} \geq y_K^{\mathcal{M}},$$

where $x_K^{\mathcal{M}}$ and $y_K^{\mathcal{M}}$ are options x and y as conceptualized in context K. Given the model's context-independence (in both senses), $x_K^{\mathcal{M}}$ and $y_K^{\mathcal{M}}$ do not depend on K (see Section 2.5). Thus, $\succsim_K^{\mathcal{M}}$ does not depend on K; we can write it as $\succsim^{\mathcal{M}}$. Therefore the choice function C is rationalizable in the classical sense by a binary relation (i.e., $\succsim^{\mathcal{M}}$).

Step 2. Now assume Axioms 1*, 2* and 3*. Let $\succsim^*$ be the classical revealed preference relation on X: i.e., for all options $x, y \in X$, let '$x \succsim^* y$' mean that x is chosen weakly over y in some context. We prove that C is reason-based rationalizable (for instance) by the model with context-invariant and context-unregarding motivation $\mathcal{M} = (M, \geq)$ defined as follows:

- M is the set $\mathcal{P}_{\text{option}}$ of *all* option properties.

- For all property bundles $S, T \subseteq \mathcal{P}$, '$S \geq T$' means that $x \succsim^* y$ for some options $x, y \in X$ such that $\mathcal{P}(x) = S$ and $\mathcal{P}(y) = T$.

Under this model, the options are conceptualized as follows:

$$x_K^{\mathcal{M}} = \mathcal{P}(x, K) \cap M = \mathcal{P}(x) \text{ for all } x \in X \text{ and } K \in \mathcal{K}. \tag{2}$$

Clearly, these options-as-conceptualized do not depend on the context K; hence, the induced preference relation $\succsim^{\mathcal{M}}$ $(= \succsim_K^{\mathcal{M}})$ does not depend on the context either.

Let $\succsim^{**}$ be the binary relation defined as

$$x \succsim^{**} y \Leftrightarrow [x \succsim^* y \text{ or } \mathcal{P}(x) = \mathcal{P}(y)] \text{ for all } x, y \in X.$$

We have to prove that $C = C^{\mathcal{M}}$. This follows from three facts:

(i) $C^{\mathcal{M}}$ is (classically) rationalized by $\succsim^{\mathcal{M}}$;

(ii) C is (classically) rationalized by $\succsim^*$ and by $\succsim^{**}$ (and thus, by any relation $\succsim$ such that $\succsim^* \subseteq \succsim \subseteq \succsim^{**}$);

(iii) $\succsim^* \subseteq \succsim^{\mathcal{M}} \subseteq \succsim^{**}$.

Fact (i): This holds by definition of $C^{\mathcal{M}}$.

Fact (ii): By Remark 1 (Richter's result), Axiom 3* implies that C is (classically) rationalizable by a binary relation. One of these rationalizations (in fact, the minimal one) is the classical revealed preference relation $\succsim^*$, as is easily checked and well-known (see also Richter 1971). Also, $\succsim^{**}$ rationalizes C, which can be shown as follows. Consider a context K. We have to show that

$$C(K) = \{x \in K : x \succsim^{**} y \text{ for all } y \in K\}.$$

Since $\succsim^{**}$ extends $\succsim^*$, $C(K) \subseteq \{x \in K : x \succsim^{**} y \text{ for all } y \in K\}$. Conversely, suppose $x \in K$ such that $x \succsim^{**} y$ for all $y \in K$. We show that $x \in C(K)$. If $\mathcal{P}(z) = \mathcal{P}(x)$ for all $z \in K$, then $C(K) = K$ by Axiom 1* and the fact that $C(K) \neq \varnothing$. Thus $x \in C(K)$, as required. Now let $z \in K$ such that $\mathcal{P}(z) \neq \mathcal{P}(x)$. Consider any $y \in K$. We have to show that $x \succsim^* y$. If $\mathcal{P}(y) \neq \mathcal{P}(x)$, this holds by the definition of $\succsim^{**}$ and the fact that $x \succsim^{**} y$. Now suppose $\mathcal{P}(y) = \mathcal{P}(x)$. Note that $x \succsim^* z$ (since $x \succsim^{**} z$ and $\mathcal{P}(z) \neq \mathcal{P}(x)$). So, there is a context $\widetilde{K} \in \mathcal{K}$ such that $x \in C(\widetilde{K})$. Since $\mathcal{P}(y) = \mathcal{P}(x)$ and since $\mathcal{K}$ is closed under cloning, there is a context $K' \in \mathcal{K}$ such that $K' = \widetilde{K} \cup \{y\}$. By Axiom 2* and the fact that $\{\mathcal{P}(v) : v \in \widetilde{K}\} = \{\mathcal{P}(v) : v \in K'\}$ and $x \in C(\widetilde{K})$, we have $v \in C(K')$ for some $v \in K'$ such that $\mathcal{P}(v) = \mathcal{P}(x)$. So, by Axiom 1*, $x \in C(K')$. As $x \in C(K')$ and $y \in K'$, we have $x \succsim^* y$, as required.

Fact (iii): Consider any $x, y \in X$. We have to show that

$$[x \succsim^* y \Rightarrow x \succsim^{\mathcal{M}} y] \text{ and } [x \succsim^{\mathcal{M}} y \Rightarrow x \succsim^{**} y].$$

Given that the options-as-conceptualized take the form (2), we have $x \succsim^{\mathcal{M}} y \Leftrightarrow \mathcal{P}(x) \geq \mathcal{P}(y)$. Therefore, we have to prove that

$$[x \succsim^* y \Rightarrow \mathcal{P}(x) \geq \mathcal{P}(y)] \text{ and } [\mathcal{P}(x) \geq \mathcal{P}(y) \Rightarrow x \succsim^{**} y].$$

The first of these two implications holds immediately by the definition of $\geq$. As for the second implication, we suppose $\mathcal{P}(x) \geq \mathcal{P}(y)$ and claim that $x \succsim^{**} y$. If $\mathcal{P}(x) = \mathcal{P}(y)$, the claim holds by the definition of $\succsim^{**}$. From now on, suppose $\mathcal{P}(x) \neq \mathcal{P}(y)$. Since $\mathcal{P}(x) \geq \mathcal{P}(y)$, there exist $x', y' \in X$ such that $\mathcal{P}(x') = \mathcal{P}(x)$, $\mathcal{P}(y') = \mathcal{P}(y)$ and $x' \succsim^* y'$. Since $x' \succsim^* y'$, there is a context $K \in \mathcal{K}$ such that $x' \in C(K)$ and $y' \in K$. Relying twice on the fact that $\mathcal{K}$ is closed under cloning, we can choose a context $K' \in \mathcal{K}$ such that $K' = K \cup \{x, y\}$. By Axiom 2* and the fact that $\{\mathcal{P}(z) : z \in K\} = \{\mathcal{P}(z) : z \in K'\}$ and $x' \in C(K)$, we have $v \in C(K')$ for some $v \in K'$ such that $\mathcal{P}(v) = \mathcal{P}(x')$. So, by Axiom 1*, $x \in C(K')$. Since $x \in C(K')$ and $y \in K'$, we have $x \succsim^* y$. Hence, $x \succsim^{**} y$, as required. ■

Proof of Proposition 1. Consider any reason-based model $\mathcal{M} = (M, \geq)$. Define a reason-based model with context-invariant motivation $\mathcal{M}' = (M', \geq')$ as follows:

- M' is any property set such that $M' \supseteq \cup_{K \in \mathcal{K}}(M_K \cup \mathcal{P}(K))$ $(= (\cup_{K \in \mathcal{K}} M_K) \cup \mathcal{P}_{\text{context}})$, for instance $M' = \mathcal{P}$;
- for any property bundles S, T, '$S \geq' T$' is defined to mean that there exists a context $K \in \mathcal{K}$ such that $\mathcal{P}(K) = S \cap \mathcal{P}_{\text{context}} = T \cap \mathcal{P}_{\text{context}}$ and $S \cap M_K \geq T \cap M_K$.

We prove that $C^{\mathcal{M}} = C^{\mathcal{M}'}$. Consider an arbitrary context $K \in \mathcal{K}$; we have to show that $C^{\mathcal{M}}(K) = C^{\mathcal{M}'}(K)$. We do so by proving that $\mathcal{M}$ and $\mathcal{M}'$ induce the same preference relation on X in context K. Fix options $x, y \in X$. We have to show that $x \succsim_K^{\mathcal{M}} y \Leftrightarrow x \succsim_K^{\mathcal{M}'} y$, i.e., writing $S = \mathcal{P}(x, K)$ and $T = \mathcal{P}(y, X)$, that

$$S \cap M_K \geq T \cap M_K \Leftrightarrow S \cap M' \geq' T \cap M'.$$

We will draw on the fact that (*) $\mathcal{P}(K) = S \cap \mathcal{P}_{\text{context}} = T \cap \mathcal{P}_{\text{context}}$.

'$\Rightarrow$': If $S \cap M_K \geq T \cap M_K$, then $S \geq' T$ by (*) and the definition of $\geq'$, and hence, $S \cap M' \geq' T \cap M'$.

'$\Leftarrow$': Now suppose $S \cap M' \geq' T \cap M'$. By definition of $\geq'$, there is a context $K' \in \mathcal{K}$ such that $\mathcal{P}(K') = S \cap \mathcal{P}_{\text{context}} = T \cap \mathcal{P}_{\text{context}}$ and $(S \cap M') \cap M_{K'} \geq (T \cap M') \cap M_{K'}$. We

deduce two facts: first, $\mathcal{P}(K') = \mathcal{P}(K)$ (where we use (*)); second, $S \cap M_{K'} \geq T \cap M_{K'}$ (where we use that $M_{K'} \subseteq M'$). The first fact implies that $M_{K'} = M_K$ (by the definition of a reason-based model). This, together with the second fact, implies that $S \cap M_K \geq T \cap M_K$, as required. ■

Before proving Theorem 3, we first show that Axioms 1 and 3 can be jointly summarized in the following axiom:

Axiom 3$^+$. For every option x in a context $K \in \mathcal{K}$, if the property bundle $\mathcal{P}(x, K)$ is revealed weakly preferred to the property bundle $\mathcal{P}(y, K)$ for every option y in K, then $x \in C(K)$.

Lemma 4 *Axioms 1 and 3 are jointly equivalent to Axiom 3$^+$.*

Proof. '$\Leftarrow$': First assume Axioms 1 and 3. As in Axiom 3$^+$, consider $K \in \mathcal{K}$ and $x \in K$ such that $\mathcal{P}(x, K) \succsim^C \mathcal{P}(y, K)$ for all $y \in K$. By Axiom 3, $\mathcal{P}(x, K)$ is chosen in context K. So, $C(K)$ contains some x' such that $\mathcal{P}(x', K) = \mathcal{P}(x, K)$. Hence, by Axiom 1, $x \in C(K)$.

'$\Rightarrow$': Now assume Axiom 3$^+$. Axiom 3 holds obviously. As for Axiom 1, consider $K \in \mathcal{K}$ and $x, y \in K$ such that $\mathcal{P}(x, K) = \mathcal{P}(y, K)$. We only show that $x \in C(K) \Rightarrow y \in C(K)$; the converse implication is analogous. Let $x \in C(K)$. Clearly, the property bundle $\mathcal{P}(x, K)$ is revealed weakly preferred to each feasible property bundle in this context K. The same is therefore true of the property bundle $\mathcal{P}(y, K)$ $(= \mathcal{P}(x, K))$. So, by Axiom 3$^+$, $y \in C(K)$. ■

Proof of Theorem 3. Step 1. Suppose a reason-based model with context-invariant motivation, $(M, \geq)$, rationalizes C. Axiom 1 holds obviously. To prove Axiom 3, consider a context $K \in \mathcal{K}$ and a bundle $S \subseteq \mathcal{P}$ feasible in K such that $S \succsim^C \mathcal{P}(y, K)$ for each y in K. Choose an x in K such that $S = \mathcal{P}(x, K)$. It suffices to show that $x \in C(K)$, i.e., since $(M, \geq)$ rationalizes C, that

$$\mathcal{P}(x, K) \cap M \geq \mathcal{P}(y, K) \cap M \tag{3}$$

for all $y \in K$. Consider any $y \in K$. Since $\mathcal{P}(x, K) \succsim^C \mathcal{P}(y, K)$, there exist $K' \in \mathcal{K}$ and $x', y' \in K'$ (which may depend on y) such that (i) $\mathcal{P}(x', K') = \mathcal{P}(x, K)$ and $\mathcal{P}(y', K') = \mathcal{P}(y, K)$, and (ii) $C(K') = x'$. By (ii) and the fact that $(M, \geq)$ rationalizes C,

$$\mathcal{P}(x', K') \cap M \geq \mathcal{P}(y', K') \cap M.$$

By (i), this implies (3), as required.

Step 2. Now assume Axioms 1 and 3. We show that C is rationalizable for instance by the (rather special) reason-based model with context-invariant motivation $(M, \geq) = (\mathcal{P}, \succsim^C)$, where M (which is constant) contains *all* properties, and $\geq$ is simply the relation of revealed weak preference. To show this, consider any context $K \in \mathcal{K}$ and option $x \in K$. We have to show that

$$x \in C(K) \Leftrightarrow [\mathcal{P}(x, K) \cap M \geq \mathcal{P}(y, K) \cap M \text{ for all } y \in K,$$

or equivalently, given our special definitions of M and $\geq$, that

$$x \in C(K) \Leftrightarrow [\mathcal{P}(x, K) \succsim^C \mathcal{P}(y, K) \text{ for all } y \in K.$$

The right-hand side of this equivalence implies that $x \in C(K)$ by Axiom 3^+, where this axiom holds by Lemma 4. Conversely, if $x \in C(K)$, then the right-hand side holds by the definition of the revealed preference relation $\succsim^C$. ■

Proof of Proposition 2. Let $K \in \mathcal{K}$. Suppose $P \in \mathcal{P}$ satisfies (rev1)-(rev3) for $S, S' \subseteq \mathcal{P}$. We may assume without loss of generality that S is revealed strictly preferred to S' (rather than vice versa), since (rev1)-(rev3) remain valid if S and S' are interchanged. Define both T and T' as S. Then (rev1) implies (REV1); (rev2) implies (REV2) (noting that S is revealed weakly preferred to itself); and (rev3) implies (REV3) because, as $T = T' = S$, (REV3) requires that the pair (S, S) differs minimally from the pair (S', S), which in turn reduces to (rev3). ■

We now formally re-state and prove Lemma 2.

Lemma 2 *For all revealed comparable bundles $S, T \subseteq \mathcal{P}$ and revealed comparable bundles $S', T' \subseteq \mathcal{P}$, if*

- *$V \cap \mathcal{P}_{\text{context}}$ is the same for all $V \in \{S, T, S', T'\}$ (i.e., S, T, S' and T' are feasible in the same type of context),*
- *$S \cap M_K^C = S' \cap M_K^C$ and $T \cap M_K^C = T' \cap M_K^C$ for the contexts $K \in \mathcal{K}$ such that $\mathcal{P}(K) = V \cap \mathcal{P}_{\text{context}}$ for all $V \in \{S, T, S', T'\}$,*

then $S \succsim^C T \Leftrightarrow S' \succsim^C T'$.

Proof. As one may verify, it is sufficient to show the following condition for all contexts $K \in \mathcal{K}$ and all finite property bundles $S, S', T, T' \subseteq \mathcal{P}$.

Condition ($X_{S,T}^{S',T'}$)*:* If
($a_{S,T}^{S',T'}$) $S \succsim\precsim^C T$ and $S' \succsim\precsim^C T'$,
($b_{S,T}^{S',T'}$) $V \cap \mathcal{P}_{\text{context}} = \mathcal{P}(K)$ for all $V \in \{S, T, S', T'\}$,
($c_{S,T}^{S',T'}$) $S \cap M_K^C = S' \cap M_K^C$ and $T \cap M_K^C = T' \cap M_K^C$,
then $S \succsim^C T \Leftrightarrow S' \succsim^C T'$.

Here it was possible to restrict ourselves to *finite* property bundles since whenever one of the property bundles S, S', T, T' is infinite, condition ($a_{S,T}^{S',T'}$) cannot be met (because feasible property bundles are by assumption finite).

Fix a context $K \in \mathcal{K}$. Note that the set of properties in which two property bundles $S, S' \subseteq \mathcal{P}$ differ is the symmetric difference $S \triangle S'$. We prove that ($X_{S,T}^{S',T'}$) holds for all finite $S, T, S', T' \subseteq \mathcal{P}$, by induction on $|S \triangle S'| + |T \triangle T'|$, the total number of disagreements between S and S' and between T and T'.

Initial step. First, consider any finite $S, S', T, T' \subseteq \mathcal{P}$ such that $|S \triangle S'| + |T \triangle T'| = 0$. Then $S = S'$ and $T = T'$, so that ($X_{S,T}^{S',T'}$) holds trivially because we have $S \succsim^C T \Leftrightarrow S' \succsim^C T'$ (even without assuming ($a_{S,T}^{S',T'}$)-($c_{S,T}^{S',T'}$)).

Induction step. Consider any $m > 0$ and suppose that ($X_{S,T}^{S',T'}$) holds for any finite $S, S', T, T' \subseteq \mathcal{P}$ such that $|S \triangle S'| + |T \triangle T'| < m$. Consider finite sets $S, S', T, T' \subseteq \mathcal{P}$ such that $|S \triangle S'| + |T \triangle T'| = m$ and assume ($a_{S,T}^{S',T'}$)-($c_{S,T}^{S',T'}$) hold. To show that $S \succsim^C T \Leftrightarrow S' \succsim^C T'$, we consider two cases.

Case 1. The pair (S, T) differs minimally from (S', T'), i.e., there is no revealed comparable pair of property bundles (S'', T'') $(\neq (S, T), (S', T'))$ such that S'' is weakly between S and S' and T'' is weakly between T and T'. Suppose, for a contradiction, that $S \succsim^C T \not\Leftrightarrow S' \succsim^C T'$. Since $|S \triangle S'| + |T \triangle T'| = m > 0$, we can choose a $P \in (S \triangle S') \cup (T \triangle T')$. We prove that P is revealed motivationally salient in K, i.e., that $P \in M_K^C$; this contradicts ($c_{S,T}^{S',T'}$), completing Case 1. By definition of M_K^C, this claim follows from three facts: (REV1) holds because $S \succsim^C T \not\Leftrightarrow S' \succsim^C T'$ (where $S \succsim\precsim^C T$ and $S' \succsim\precsim^C T'$ by ($a_{S,T}^{S',T'}$)); (REV2) holds because S and S' differ in P or T and T' differ in P (since $P \in (S \triangle S') \cup (T \triangle T')$) and because of ($b_{S,T}^{S',T'}$); (REV3) holds because the pair (S, T) differs minimally from (S', T') by assumption of Case 1.

Case 2. The pair (S, T) does *not* differ minimally from (S', T'). Then we may choose a revealed comparable pair of property bundles (S'', T'') $(\neq (S, T), (S', T'))$ such that S'' is weakly between S and S' and T'' is weakly between T and T'. Observe that $|S \triangle S''| < |S \triangle S'|$ or $|T \triangle T''| < |T \triangle T'|$ (possibly both). So, $|S \triangle S''| + |T \triangle T''| < |S \triangle S'| + |T \triangle T'|$, and hence, $|S \triangle S''| + |T \triangle T''| < m$. Also, noting that S'' and T''

are finite (since they are feasible in some context as $S'' \succsim\precsim^C T''$), it follows by induction hypothesis that the implication $(\mathrm{X}_{S,T}^{S'',T''})$ holds. Now the three antecedent conditions of this implication hold. Condition $(\mathrm{b}_{S,T}^{S'',T''})$ holds because, first, $S \cap \mathcal{P}_{\text{context}} = T \cap \mathcal{P}_{\text{context}} = \mathcal{P}(K)$ by $(\mathrm{b}_{S,T}^{S',T'})$, second, $S'' \cap \mathcal{P}_{\text{context}} = S \cap \mathcal{P}_{\text{context}} = S' \cap \mathcal{P}_{\text{context}}$ by $(\mathrm{b}_{S,T}^{S',T'})$ and the fact that S'' is weakly between S and S', and, third, $T'' \cap \mathcal{P}_{\text{context}} = T \cap \mathcal{P}_{\text{context}} = T' \cap \mathcal{P}_{\text{context}}$ for analogous reasons. Condition $(\mathrm{a}_{S,T}^{S'',T''})$ follows from $(\mathrm{a}_{S,T}^{S',T'})$ and the fact that $S'' \succsim\precsim^C T''$. Condition $(\mathrm{c}_{S,T}^{S'',T''})$ may be deduced from $(\mathrm{c}_{S,T}^{S',T'})$ and the fact that S'' is weakly between S and S' and T'' is weakly between T and T'. From $(\mathrm{X}_{S,T}^{S'',T''})$ and $(\mathrm{a}_{S,T}^{S'',T''})$-$(\mathrm{c}_{S,T}^{S'',T''})$ it follows that $S \succsim^C T \Leftrightarrow S'' \succsim^C T''$.

By an analogous reasoning applied to the sets S', S'', T', T'' (rather than S, S'', T, T''), we have $S' \succsim^C T' \Leftrightarrow S'' \succsim^C T''$. This equivalence and the previous one jointly imply the equivalence $S \succsim^C T \Leftrightarrow S' \succsim^C T'$, as required. ■

Proof of Lemma 3. Axiom 1** obviously implies Axiom 1. Now assume Axiom 1. Fix a context $K \in \mathcal{K}$ and options $x, y \in K$ such that $x_K^C = y_K^C$. We only show that $x \in C(K) \Rightarrow y \in C(K)$; the converse implication ('$\Leftarrow$') holds analogously. Suppose $x \in C(K)$. The proof is in three claims (only the last of which draws on Axiom 1).

Claim 1. There exists a finite sequence $(S_1, ..., S_m)$ of property bundles such that (i) $S_1 = \mathcal{P}(x, K)$ and $S_m = \mathcal{P}(y, K)$, (ii) $S_1 \succsim\precsim^C S_j$ for each $j \in \{1, ..., m\}$, (iii) for all $j, j', j'' \in \{1, ..., m\}$, if $j \leq j' \leq j''$ then $S_{j'}$ is weakly between S_j and $S_{j''}$, (iv) for all $j \in \{1, ..., m-1\}$, $S_j \neq S_{j+1}$, and (v) for all $j \in \{1, ..., m-1\}$, no property bundle $S \subseteq \mathcal{P}$ is strictly between S_j and S_{j+1} and satisfies $S \succsim\precsim^C S_1$.

Let $\mathbf{S}$ be the set of all finite sequences $(S_1, ..., S_m)$ of property bundles satisfying the first four conditions (i)-(iv). Since $x \in C(K)$, we have $\mathcal{P}(x, K) \succsim^C \mathcal{P}(x, K)$ and $\mathcal{P}(x, K) \succsim^C \mathcal{P}(y, K)$. In particular, $\mathcal{P}(x, K) \succsim\precsim^C \mathcal{P}(x, K)$ and $\mathcal{P}(x, K) \succsim\precsim^C \mathcal{P}(y, K)$. So, $\mathbf{S}$ contains the sequence $(\mathcal{P}(x, K), \mathcal{P}(y, K))$ if $\mathcal{P}(x, K) \neq \mathcal{P}(y, K)$ and contains the single-component sequence $(\mathcal{P}(x, K))$ if $\mathcal{P}(x, K) = \mathcal{P}(y, K)$. Hence, $\mathbf{S} \neq \varnothing$.

Next, note that since the property bundles $\mathcal{P}(x, K)$ and $\mathcal{P}(y, K)$ are finite, the set $\mathcal{P}(x, K) \triangle \mathcal{P}(y, K)$ of properties in which they differ is also finite. For all $(S_1, ..., S_m) \in \mathbf{S}$ we have $m - 1 \leq |\mathcal{P}(x, K) \triangle \mathcal{P}(y, K)|$ $(= |S_1 \triangle S_m|)$. To prove this, consider any $(S_1, ..., S_m) \in \mathbf{S}$ and let us show by induction that $|S_1 \triangle S_j| \geq j - 1$ for all $j \in \{1, ..., m\}$. For $j = 1$ this is obviously true. Now consider any $j \in \{1, ..., m-1\}$ such that $|S_1 \triangle S_j| \geq j - 1$. We have $|S_1 \triangle S_{j+1}| \geq |S_1 \triangle S_j| + 1 \geq j$, where the first inequality holds because S_j is strictly between S_1 and S_{j+1} by (iii) and (iv), and the second inequality holds because $|S_1 \triangle S_j| \geq j - 1$. This completes the inductive argument.

As shown so far, $\mathbf{S}$ is non-empty and the length of sequences in $\mathbf{S}$ has a finite upper bound (given by $|\mathcal{P}(x,K) \triangle \mathcal{P}(y,K)| + 1$). So there exists a longest sequence in $\mathbf{S}$. Call it $(S_1, ..., S_m)$. We complete the proof of the claim by showing that this sequence also satisfies condition (v).

Suppose, for a contradiction, that $j \in \{1, ..., m-1\}$ and there is a property bundle $S \subseteq \mathcal{P}$ which is strictly between S_j and S_{j+1} and satisfies $S_1 \succsim\precsim^C S$. Form the augmented sequence $(S_1, ..., S_j, S, S_{j+1}, ..., S_m)$. We show that this sequence satisfies (i)-(iv), i.e., belongs to $\mathbf{S}$, a contradiction, since the sequence is longer than $(S_1, ..., S_m)$.

First, the augmented sequence obviously still satisfies (i), (ii) and (iv). It remains to show that it also satisfies (iii). To do so, we consider indices $i, i' \in \{1, ..., m\}$ and have to show three things:

(*) if $i \leq i' \leq j$, then $S_{i'}$ is (weakly) between S_i and S;

(**) if $j+1 \leq i \leq i'$, then S_i is between S and $S_{i'}$;

(***) if $i \leq j$ and $j+1 \leq i'$, then S is between S_i and $S_{i'}$.

Regarding (*), assume $i \leq i' \leq j$, and consider a $P \in \mathcal{P}$ on which S_i and S agree. We have to show that $S_{i'}$ agrees on P with S_i (and S). Since S is strictly between S_j and S_{j+1}, S agrees on P with at least one of S_j and S_{j+1}. Let $j' \in \{j, j+1\}$ be such that S and $S_{j'}$ agree on P. So, S_i and $S_{j'}$ also agree on P. Hence, since $S_{i'}$ is between S_i and $S_{j'}$ (as the original sequence $(S_1, ..., S_m)$ satisfies (iii)), $S_{i'}$ agrees on P with S_i, as required to prove (*).

The proof of (**) is analogous to that of (*).

Regarding (***), assume $i \leq j$ and $j+1 \leq i'$. Consider any $P \in \mathcal{P}$ on which S_i and $S_{i'}$ agree. We have to show that S agrees with S_i (and $S_{i'}$) on P. Since the original sequence $(S_1, ..., S_m)$ satisfies (iii), S_j is between S_i and $S_{i'}$ (if $i = j$ trivially), and so S_j and S_i agree on P. By an analogous argument, S_{j+1} and S_i agree on P. Hence, S_j and S_{j+1} agree on P. So, as S is (strictly) between S_j and S_{j+1}, S agrees on P with S_j, and hence also with S_i. This shows (***), completing the proof of Claim 1.

Claim 2: If $(S_1, ..., S_m)$ is any sequence of property bundles satisfying the conditions (i)-(v) in Claim 1, then for all $j \in \{1, ..., m\}$ neither of the bundles S_j and S_1 $(= \mathcal{P}(x,K))$ is revealed strictly preferred to the other.

The proof is by induction on j. If $j = 1$, the claim holds trivially. Now consider $j \in \{1, ..., m-1\}$ and assume neither of the sets S_j and S_1 is revealed strictly preferred to the other. Suppose, for a contradiction, that one of the sets S_{j+1} and S_1 is revealed strictly preferred to the other one. We assume without loss of generality that S_1 is revealed strictly preferred to S_{j+1} (the proof proceeds analogously in the other case).

Since $S_j \neq S_{j+1}$ by (iv), we may select a property $P \in S_j \triangle S_{j+1}$. Now P is revealed motivationally salient in K, i.e., $P \in M_K^C$. We show this by verifying the criteria (REV1)-(REV3) for the pairs of bundles $(S_j, , S_1)$ and (S_{j+1}, S_1). First, S_j is revealed weakly preferred to S_1 (because S_1 is not revealed strictly preferred to S_j by induction hypothesis and because $S_1 \succsim^C S_j$ by (ii)), while S_1 is revealed strictly preferred to S_{j+1}, where these two choices occur in contexts with the same properties as K (because $S_1 = \mathcal{P}(x, K)$ by (i)). Second, S_j and S_{j+1} differ in P (since $P \in S_j \triangle S_{j+1}$). Third, by (v) the pair (S_1, S_j) differs minimally from (S_1, S_{j+1}) in the sense defined in (REV3).

Now, since $P \in M_K^C$ and since S_j and S_{j+1} differ in P, we have $(S_j \triangle S_{j+1}) \cap M_K^C \neq \varnothing$. Further, $S_j \triangle S_{j+1} \subseteq S_1 \triangle S_m$. Indeed, if a property P does *not* belong to $S_1 \triangle S_m$, then S_1 and S_m agree on P, so that all of S_j, S_{j+1}, S_1 and S_m agree on P (since S_j and S_{j+1} are each weakly between S_1 and S_m by (iii)), which implies that P is *not* contained in $S_j \triangle S_{j+1}$. Since $(S_j \triangle S_{j+1}) \cap M_K^C \neq \varnothing$ and $S_j \triangle S_{j+1} \subseteq S_1 \triangle S_m$, we have $(S_1 \triangle S_m) \cap M_K^C \neq \varnothing$. So, $S_1 \cap M_K^C \neq S_m \cap M_K^C$. Hence, by (i), $\mathcal{P}(x, K) \cap M_K^C \neq \mathcal{P}(y, K) \cap M_K^C$, i.e., $x_K^C \neq y_K^C$, contradicting the initial assumption that $x_K^C = y_K^C$. This proves Claim 2.

Claim 3. $y \in C(K)$ (which completes the proof of Axiom 1**).

By Claims 1 and 2, $\mathcal{P}(x, K)$ is not revealed strictly preferred to $\mathcal{P}(y, K)$. So, since $\mathcal{P}(x, K)$ is chosen in K (as $x \in C(K)$), so is $\mathcal{P}(y, K)$. Using Axiom 1 it follows that $y \in C(K)$. ■

Proof of Theorem 4 (in its strengthened form given in its footnote). Assume the domain of contexts $\mathcal{K}$ is rich. We prove the necessity of the axioms (step 1), the sufficiency of the axioms (step 2), and the essential uniqueness claim (step 3).

Step 1. Suppose C has a reason-based rationalization with revealed motivation $(M^C, \geq)$. Axioms 1 and 3 hold by Theorem 1. To prove that Axiom 2** holds, consider contexts $K, K' \in \mathcal{K}$ such that (*) $\{x_K^C : x \in K\} = \{x'^C_{K'} : x' \in K'\}$. We only show that $\{y_K^C : y \in C(K)\} \subseteq \{y'^C_{K'} : y' \in C(K')\}$, since the converse inclusion holds analogously. Consider any $y \in C(K)$. We have to show that $y_K^C \in \{y'^C_{K'}(y') : y' \in C(K')\}$. By (*) there is a $y' \in K'$ such that (**) $y_K^C = y'^C_{K'}$. It remains to show that $y' \in C(K')$. Since $y \in C(K)$ and C is rationalized by $(M^C, \geq)$, we have

$$y_K^C \geq x_K^C \text{ for all } x \in K.$$

By (*) and (**), this implies that

$$y'^C_{K'} \geq x'^C_{K'} \text{ for all } x' \in K'.$$

It follows that $y' \in C(K')$, again because C is rationalized by $(M^C, \geq)$. This proves Axiom 2**.

Step 2. Conversely, assume Axioms 1, 2** and 3. We show in two claims that the revealed model $(M^C, \geq^C)$ rationalizes C.

Claim 1. For all contexts $K, K' \in \mathcal{K}$ and all options $x, y \in K$ and $x', y' \in K'$, if $x^C_K = x'^C_{K'}$ and $y^C_K = y'^C_{K'}$ then $\mathcal{P}(x, K) \succsim^C \mathcal{P}(y, K) \Leftrightarrow \mathcal{P}(x', K') \succsim^C \mathcal{P}(y', K')$.

Consider $K, K' \in \mathcal{K}$, $x, y \in K$ and $x', y' \in K'$ such that $x^C_K = x'^C_{K'}$ and $y^C_K = y'^C_{K'}$. We assume that $\mathcal{P}(x', K') \succsim^C \mathcal{P}(y', K')$ and show that $\mathcal{P}(x, K) \succsim^C \mathcal{P}(y, K)$; the converse implication is analogous.

Since $\mathcal{K}$ is rich and the bundles $\mathcal{P}(x, K)$ and $\mathcal{P}(y, K)$ are feasible in K, there is a context $L \in \mathcal{K}$ in which they are the only feasible bundles:

$$\{\mathcal{P}(z, L) : z \in L\} = \{\mathcal{P}(x, K), \mathcal{P}(y, K)\}. \tag{4}$$

Now $\mathcal{P}(K) = \mathcal{P}(L)$, since each side of this equality can be written as $S \cap \mathcal{P}_{\text{context}}$ for a bundle S feasible in K and L. It follows that $M^C_K = M^C_L$. On each side of (4) we now intersect each contained bundle with M^C_K $(= M^C_L)$. This yields a new identity:

$$\{z^C_L : z \in L\} = \{x^C_K, y^C_K\}. \tag{5}$$

The steps taken for x, y, K are now repeated for x', y', K'. By the richness of $\mathcal{K}$ and the feasibility of the bundles $\mathcal{P}(x', K')$ and $\mathcal{P}(y', K')$ in K', there is a context $L' \in \mathcal{K}$ such that

$$\{\mathcal{P}(z, L') : z \in L'\} = \{\mathcal{P}(x', K'), \mathcal{P}(y', K')\}. \tag{6}$$

By arguments made similarly above, it follows that $M^C_{K'} = M^C_{L'}$ and

$$\{z^C_{L'} : z \in L'\} = \{x'^C_{K'}, y'^C_{K'}\}. \tag{7}$$

From (5), (7) and the assumption that $x^C_K = x'^C_{K'}$ and $y^C_K = y'^C_{K'}$, we deduce that $\{z^C_L : z \in L\} = \{z^C_{L'} : z \in L'\}$. So, by Axiom 2**,

$$\{z^C_L : z \in C(L)\} = \{z^C_{L'} : z \in C(L')\}. \tag{8}$$

By Axiom 3, (6) and the assumption that the bundle $\mathcal{P}(x', K')$ is revealed weakly preferred to $\mathcal{P}(y', K')$ (and thus also to itself), the bundle $\mathcal{P}(x', K')$ is chosen in L':

$$\mathcal{P}(x', K') \in \{\mathcal{P}(z, L') : z \in C(L')\}.$$

Intersecting on both sides of this relation with $M^C_{K'}$ $(= M^C_{L'})$ yields

$$x'^C_{K'} \in \{z^C_{L'} : z \in C(L')\}.$$

By (8) and the fact that $x^C_K = x'^C_{K'}$, we can rewrite the last relation as

$$x^C_K \in \{z^C_L : z \in C(L)\}.$$

Pick a $z \in C(L)$ such that $x^C_K = z^C_L$. By (4) we can also pick a $w \in L$ such that $\mathcal{P}(w, L) = \mathcal{P}(x, K)$. Intersecting each side of this equation with M^C_L $(= M^C_K)$ yields $w^C_L = x^C_K$. Hence, $w^C_L = z^C_L$. By Axiom 1** (which holds by Lemma 3), it follows that $w \in C(L)$. So, the bundle $\mathcal{P}(w, L) = \mathcal{P}(x, K)$ is revealed weakly preferred to any bundle feasible in L, hence by (4) to $\mathcal{P}(y, K)$.

Claim 2. $(M^C, \geq^C)$ rationalizes C (which completes the sufficiency proof).

We consider any $K \in \mathcal{K}$ and $x \in K$ and have to show that

$$x \in C(K) \Leftrightarrow \left[x^C_K \geq^C y^C_K \text{ for all } y \in K\right].$$

First, if $x \in C(K)$, then for all $y \in K$ we indeed have $x^C_K(x) \geq^C y^C_K$, immediately by definition of $\geq^C$. Now assume that $x^C_K \geq^C y^C_K$ for all $y \in K$. Consider any $y \in K$. Since $x^C_K \geq^C y^C_K$, by definition of $\geq^C$ there exist $K' \in \mathcal{K}$ and $x', y' \in K'$ (all of which may depend on y) such that $x^C_K = x'^C_{K'}$, $y^C_K = y'^C_{K'}$ and $x' \in C(K')$. Since $x' \in C(K')$ and $y' \in K'$, we have $\mathcal{P}(x', K') \succsim^C \mathcal{P}(y', K')$. So, by Claim 1, $\mathcal{P}(x, K) \succsim^C \mathcal{P}(y, K)$. Since this is true for all $y \in K$, we have $x \in C(K)$, by Axiom 3^+ (which holds by Lemma 4).

Step 3. We now consider an arbitrary rationalization of C with revealed motivation $(M^C, \geq)$, and have to show that it is essentially identical to $(M^C, \geq^C)$ (which rationalizes C by part 2). As the two models ascribe the same motivation to the agent, it remains to show that $\geq$ and $\geq^C$ coincide wherever they are choice-behaviourally relevant. Consider a pair of bundles $S, T \subseteq \mathcal{P}$ at which $\geq$ and $\geq^C$ are choice-behaviourally relevant; we can thus pick $K^* \in \mathcal{K}$ and $x, y \in K^*$ such that

$$S = x^C_{K^*} \text{ and } T = y^C_{K^*}. \tag{9}$$

We have to show that $S \geq T \Leftrightarrow S \geq^C T$.

Claim 3. $S \geq S$ and $S \geq^C S$.

Since the domain of contexts $\mathcal{K}$ is rich and the (identical) property bundles $\mathcal{P}(x, K^*)$ and $\mathcal{P}(x, K^*)$ are feasible in context K^*, there is a context $\overline{K} \in \mathcal{K}$ in which only the bundle $\mathcal{P}(x, K^*)$ is feasible. Choose any $\overline{x} \in C(\overline{K})$. Clearly,

$$\mathcal{P}(x, K^*) = \mathcal{P}(\overline{x}, \overline{K}). \tag{10}$$

It follows that $\mathcal{P}(K^*) = \mathcal{P}(\overline{K})$, and hence, that $M^C_{K^*} = M^C_{\overline{K}}$. This and (10) imply that $x^C_{K^*} = \overline{x}^C_{\overline{K}}$. Now since $\overline{x} \in C(\overline{K})$ and each of the models $(M^C, \geq)$ and $(M^C, \geq^C)$ rationalizes C, we have $\overline{x}^C_{\overline{K}} \geq \overline{x}^C_{\overline{K}}$ and $\overline{x}^C_{\overline{K}} \geq^C \overline{x}^C_{\overline{K}}$. This proves the claim since $S = x^C_{K^*} = \overline{x}^C_{\overline{K}}$.

Claim 4. $S \geq T \Leftrightarrow S \geq^C T$ (which completes the proof).

Since $\mathcal{P}(x, K^*)$ and $\mathcal{P}(y, K^*)$ are both feasible in K^*, richness of $\mathcal{K}$ implies that there is a context $\tilde{K} \in \mathcal{K}$ in which only these two property bundles are feasible. Choose any $\tilde{x}, \tilde{y} \in \tilde{K}$ such that

$$\mathcal{P}(\tilde{x}, \tilde{K}) = \mathcal{P}(x, K^*) \text{ and } \mathcal{P}(\tilde{y}, \tilde{K}) = \mathcal{P}(y, K^*). \tag{11}$$

From any of these equations it follows that $\mathcal{P}(\tilde{K}) = \mathcal{P}(K^*)$, whence $M^C_{\tilde{K}} = M^C_{K^*}$. This and the equations (9) and (11) imply that $S = \tilde{x}^C_{\tilde{K}}$ and $T = \tilde{y}^C_{\tilde{K}}$. So, $\{z^C_{\tilde{K}} : z \in \tilde{K}\} = \{S, T\}$. Hence, as the model $(M^C, \geq)$ rationalizes C and as $S \geq S$ by Claim 3,

$$\tilde{x} \in C(\tilde{K}) \Leftrightarrow S \geq T. \tag{12}$$

Analogously, as $(M^C, \geq^C)$ rationalizes C and as $S \geq^C S$,

$$\tilde{x} \in C(\tilde{K}) \Leftrightarrow S \geq^C T. \tag{13}$$

The equivalences (12) and (13) imply that $S \geq T \Leftrightarrow S \geq^C T$. ■

Proof of Proposition 3. Let $\mathcal{M}' = (M', \geq)$ be a reason-based model for a domain $\mathcal{D} \subseteq \mathcal{K}^+$. Regarding part (a), if all M'_K coincide, then obviously $CAU^{\mathcal{M}'} = \varnothing$; and if $CAU^{\mathcal{M}'} = \varnothing$, then part (b) will imply that all M'_K coincide. It thus remains to prove part (b). We proceed by contraposition. Let $K, K' \in \mathcal{D}$ satisfy $M'_K \neq M'_{K'}$. Since $\mathcal{P}(x, K)$ and $\mathcal{P}(x, K')$ are finite for any $x \in X$, $\mathcal{P}(K)$ and $\mathcal{P}(K')$ are finite, and thus the 'disagreement set' $\mathcal{P}(K) \triangle \mathcal{P}(K')$ is finite. So, as one easily checks, there is a finite sequence $K_1, ..., K_n \in \mathcal{D}$ with $K_1 = K$, $K_n = K'$ such that for each $m \in \{1, ..., n-1\}$ the contexts K_m and K_{m+1} differ minimally (in the sense of (cau3)). Since $M'_{K_1} \neq M'_{K_n}$, there is an $m \in \{1, ..., n-1\}$ such that $M'_{K_m} \neq M'_{K_{m+1}}$. By the definition of reason-based models, it follows that $\mathcal{P}(K_m) \neq \mathcal{P}(K_{m+1})$. Hence we may pick a context property $P \in \mathcal{P}(K_m) \triangle \mathcal{P}(K_{m+1})$. It follows that $P \in \mathcal{P}(K) \triangle \mathcal{P}(K')$. So, since also $P \in CAU^{\mathcal{M}'}$ (as the criteria (cau1)-(cau3) hold for the contexts K_m and K_{m+1}), we have $P \in (\mathcal{P}(K) \cap CAU^{\mathcal{M}'}) \triangle (\mathcal{P}(K') \cap CAU^{\mathcal{M}'})$. Hence, $\mathcal{P}(K) \cap CAU^{\mathcal{M}'} \neq \mathcal{P}(K') \cap CAU^{\mathcal{M}'}$. ■

Proof of Remark 2. Consider a rationalization $\mathcal{M} = (M, \geq)$ of the choice function C. Let $\mathcal{M}^1 = (M^1, \geq)$, $\mathcal{M}^2 = (M^2, \geq)$, and $\mathcal{M}^3 = (M^3, \geq)$ be the models used

to define, respectively, the cautious, semi-courageous, and courageous predictors, with corresponding domains $\mathcal{D}^1$, $\mathcal{D}^2$, and $\mathcal{D}^3$.

(a) $C^{\mathcal{M}^1}$ extends C because $\mathcal{M}^1$ extends $\mathcal{M}$ (as a consequence of the definition of $\mathcal{M}^1$) and $C^{\mathcal{M}} = C$ (by assumption).

(b) We prove that $C^{\mathcal{M}^2}$ extends $C^{\mathcal{M}^1}$ by showing that $\mathcal{M}^2$ extends $\mathcal{M}^1$. Consider any $K \in \mathcal{D}^1$. We have to show that $K \in \mathcal{D}^2$ and $M^1_K = M^2_K$. Since $K \in \mathcal{D}^1$ there is an $L \in \mathcal{K}$ such that $\{P(x,K) : x \in K\} = \{P(x,L) : x \in L\}$. One easily verifies the conditions (i) (by using the same context L) and (ii) (by using the context $L' := L$).

(c) It suffices to show that $\mathcal{M}^3$ extends $\mathcal{M}^2$. Let $K \in \mathcal{D}^2$; so conditions (i) and (ii) hold. We have to show that $K \in \mathcal{D}^3$ and $M^2_K = M^3_K$. Now (i) immediately implies (i*) (use the same $L \in \mathcal{K}$), and so $K \in \mathcal{D}^3$. Moreover, $M^2_K = M^3_K$, because each side equals M_L for L as in (i). ■

Proof of Theorem 5. Consider a rationalization $\mathcal{M} = (M, \geq)$ of the choice function C. We use the notation from our proof of Remark 2. Further, for any model $\mathcal{M}'$, the set of feasible options *as conceptualized in a context* K (from the domain of $\mathcal{M}'$) is denoted $K^{\mathcal{M}'} := \{x^{\mathcal{M}'}_K : x \in K\}$.

(a) Suppose C^+ is rationalizable by an arbitrary model $\mathcal{M}^+ = (M^+, \geq^+)$ on the domain $\mathcal{K}^+$. Consider any $K \in \mathcal{D}^1$ and $x \in K$. We have to show that $x \in C^{\mathcal{M}^1}(K) \Leftrightarrow x \in C^+(K)$. As $K \in \mathcal{D}^1$ we can pick an $L \in \mathcal{K}$ such that

$$\{\mathcal{P}(y,K) : y \in K\} = \{\mathcal{P}(y,L) : y \in L\}. \tag{14}$$

So $K^{\mathcal{M}^1} = L^{\mathcal{M}^1}$ and $K^{\mathcal{M}^+} = L^{\mathcal{M}^+}$ (though perhaps $K^{\mathcal{M}^1} \neq K^{\mathcal{M}^+}$). Now pick a $z \in L$ such that $\mathcal{P}(x,K) = \mathcal{P}(z,L)$ (which is possible by (14)). It follows that $x^{\mathcal{M}^1}_K = z^{\mathcal{M}^1}_L$ and $x^{\mathcal{M}^+}_K = z^{\mathcal{M}^+}_L$. We show the claimed equivalence by proving that each side holds if and only if $z \in C(L)$:

$x \in C^{\mathcal{M}^1}(K)$	$\Leftrightarrow x^{\mathcal{M}^1}_K \geq S$ for all $S \in K^{\mathcal{M}^1}$	by definition of $C^{\mathcal{M}^1}$
	$\Leftrightarrow z^{\mathcal{M}^1}_L \geq S$ for all $S \in L^{\mathcal{M}^1}$	as $x^{\mathcal{M}^1}_K = z^{\mathcal{M}^1}_L$ and $K^{\mathcal{M}^1} = L^{\mathcal{M}^1}$
	$\Leftrightarrow z \in C^{\mathcal{M}^1}(L)$	by definition of $C^{\mathcal{M}^1}$
	$\Leftrightarrow z \in C(L)$	as $C^{\mathcal{M}^1}(L) = C(L)$ by Remark 2,
$x \in C^+(K)$	$\Leftrightarrow x \in C^{\mathcal{M}^+}(K)$	as $C^{\mathcal{M}^+} = C^+$
	$\Leftrightarrow x^{\mathcal{M}^+}_K \geq^+ S$ for all $S \in K^{\mathcal{M}^+}$	by definition of $C^{\mathcal{M}^+}$
	$\Leftrightarrow z^{\mathcal{M}^+}_L \geq^+ S$ for all $S \in L^{\mathcal{M}^+}$	as $x^{\mathcal{M}^+}_K = z^{\mathcal{M}^+}_L$ and $K^{\mathcal{M}^+} = L^{\mathcal{M}^+}$
	$\Leftrightarrow z \in C^{\mathcal{M}^+}(L)$	by definition of $C^{\mathcal{M}^+}$
	$\Leftrightarrow z \in C(L)$	as $C^{\mathcal{M}^+}(L) = C^+(L) = C(L)$.

(b) Now let C^+ be rationalizable by an extension $\mathcal{M}^+ = (M^+, \geq)$ of $\mathcal{M}$. Let $K \in \mathcal{D}^2$ and $x \in K$. We show that $x \in C^{\mathcal{M}^2}(K) \Leftrightarrow x \in C^+(K)$. As $K \in \mathcal{D}^2$ we can pick $L, L' \in \mathcal{K}$ such that $\mathcal{P}(L) = \mathcal{P}(K)$ and (*) $K^{\mathcal{M}^2} = (L')^{\mathcal{M}^2}$. By (*) we can choose a $z \in L'$ such that (**) $x_K^{\mathcal{M}^2} = z_{L'}^{\mathcal{M}^2}$. Since $M_{L'}^+ = M_{L'}^2$ $(= M_{L'})$,

$$(L')^{\mathcal{M}^+} = (L')^{\mathcal{M}^2} \text{ and } z_{L'}^{\mathcal{M}^+} = z_{L'}^{\mathcal{M}^2}. \tag{15}$$

As $M_L^+ = M_L^2$ $(= M_L)$ and $\mathcal{P}(L) = \mathcal{P}(K)$, we have $M_K^+ = M_K^2$, and thus

$$K^{\mathcal{M}^+} = K^{\mathcal{M}^2} \text{ and } x_K^{\mathcal{M}^+} = x_K^{\mathcal{M}^2}. \tag{16}$$

By (*), (**), (15) and (16), we have (***) $K^{\mathcal{M}^+} = (L')^{\mathcal{M}^+}$ and (****) $x_K^{\mathcal{M}^+} = z_{L'}^{\mathcal{M}^+}$.

One can show the claimed equivalence by proving that each side holds if and only if $z \in C(L)$. One should follow the steps taken similarly in the proof of part (a): it suffices to replace L by L' and $\mathcal{M}^1$ by $\mathcal{M}^2$, and to apply the identities (*)-(****).

(c) Finally, let C^+ be rationalizable by an extension $\mathcal{M}^+ = (M^+, \geq)$ of $\mathcal{M}$ with $CAU^{\mathcal{M}^+} = CAU^{\mathcal{M}}$. Let $K \in \mathcal{D}^3$ and $x \in K$. We prove $x \in C^{\mathcal{M}^3}(K) \Leftrightarrow x \in C^+(K)$. Since $K \in \mathcal{D}^3$, we can pick $L, L' \in \mathcal{K}$ such that $\mathcal{P}(L) \cap CAU^{\mathcal{M}} = \mathcal{P}(K) \cap CAU^{\mathcal{M}}$, $M_K^3 = M_L^3$, and ($^+$) $K^{\mathcal{M}^3} = (L')^{\mathcal{M}^3}$. Since $CAU^{\mathcal{M}^+} = CAU^{\mathcal{M}}$ and $\mathcal{P}(L) \cap CAU^{\mathcal{M}} = \mathcal{P}(K) \cap CAU^{\mathcal{M}}$, we have $\mathcal{P}(L) \cap CAU^{\mathcal{M}^+} = \mathcal{P}(K) \cap CAU^{\mathcal{M}^+}$, and thus by Proposition 3 $M_L^+ = M_K^+$. By ($^+$) there is a $z \in L'$ such that ($^{++}$) $x_K^{\mathcal{M}^3} = z_{L'}^{\mathcal{M}^3}$. Since $M_{L'}^+ = M_{L'}^3$ $(= M_{L'})$, we have

$$(L')^{\mathcal{M}^+} = (L')^{\mathcal{M}^3} \text{ and } z_{L'}^{\mathcal{M}^+} = z_{L'}^{\mathcal{M}^3}. \tag{17}$$

Since $M_L^+ = M_L^3$ $(= M_L)$, $M_L^+ = M_K^+$ and $M_L^3 = M_K^3$, we have $M_K^+ = M_K^3$, and thus

$$K^{\mathcal{M}^+} = K^{\mathcal{M}^3} \text{ and } x_K^{\mathcal{M}^+} = x_K^{\mathcal{M}^3}. \tag{18}$$

By ($^+$), ($^{++}$), (17) and (18), we have ($^{+++}$) $K^{\mathcal{M}^+} = (L')^{\mathcal{M}^+}$ and ($^{++++}$) $x_K^{\mathcal{M}^+} = z_{L'}^{\mathcal{M}^+}$. The claimed equivalence can once again be proved by establishing that each side holds if and only if $z \in C(L)$; one should use the same argument as for part (a), replacing L by L' and $\mathcal{M}^1$ by $\mathcal{M}^3$, and drawing on the identities ($^+$)-($^{++++}$). ■