



Universiteit
Leiden
The Netherlands

Better predictions when models are wrong or underspecified

Ommen, M. van

Citation

Ommen, M. van. (2015, June 10). *Better predictions when models are wrong or underspecified*. Retrieved from <https://hdl.handle.net/1887/33204>

Version: Corrected Publisher's Version

License: [Licence agreement concerning inclusion of doctoral thesis in the Institutional Repository of the University of Leiden](#)

Downloaded from: <https://hdl.handle.net/1887/33204>

Note: To cite this publication please use the final published version (if applicable).

Cover Page



Universiteit Leiden



The handle <http://hdl.handle.net/1887/33204> holds various files of this Leiden University dissertation

Author: Ommen, Thijs van

Title: Better predictions when models are wrong or underspecified

Issue Date: 2015-06-10

Chapter 7

Properties of Message Structures in Probability Updating Games

This chapter continues the analysis of probability updating games, which were introduced in Chapter 6. Familiarity with parts of that chapter will be assumed here. In particular, Section 6.2 introduced the problem statement and terminology, and Section 6.5.2 discussed RCAR strategies and RCAR vectors, which will play a central role in this chapter.

7.1 Introduction

Our results in Chapter 6 have focused on properties of the loss function L . However, the characterization theorems in the previous chapter tell us how to *recognize* worst-case optimal strategies, but not how to *find* them efficiently. To progress with this task, we also need to understand a game's message structure $\mathcal{Y}$. That is the motivation behind this chapter. Though the task remains hard to solve in general (for example, if the messages in $\mathcal{Y}$ have different sizes, we cannot do much unless we can reduce the game to a simpler one), we find several interesting results for specific cases.

As we saw in the previous chapter, the loss function and its properties play a large role in the study of worst-case optimal probability updating strategies. In particular, different strategies will in general be worst-case optimal under different loss functions. This is very different for situations where our uncertainty is expressed by a single distribution rather than a set of distributions (the possible quizmaster strategies / coarsening mechanisms). In those situations, the only rational approach to probability updating is naive conditioning, which requires just the original distribution (p) and the message $y \subset \mathcal{X}$ to compute $P(X = x \mid X \in y)$. If however $\mathcal{Y}$ is not a partition of $\mathcal{X}$, then

our uncertainty is expressed by a set of many distributions, and in general we also need to know $\mathcal{Y}$ and L to determine worst-case optimal strategies in our games. However, one of our main results in this chapter shows that for certain classes of message structures, the choice of loss function does not affect the quizmaster's worst-case optimal strategy. In these situations, the procedure of worst-case optimal probability updating becomes more similar to that of naive conditioning, because now it suffices to know just $\mathcal{Y}$ on top of what naive conditioning requires. For these message structures, the distributions of outcomes given messages we derive are in a more general sense 'optimal', expressing what a cautious experimenter *should* believe after receiving new data.

We first show a simple method of simplifying message structures in Section 7.2; there we will also see that if $\mathcal{Y}$ is a partition of $\mathcal{X}$, naive conditioning is worst-case optimal. In Section 7.3, we consider symmetry properties that worst-case optimal strategies must have, provided that the loss function also obeys a form of symmetry defined in Section 7.3.1. Then in Section 7.4, we show two classes of message structures for which the worst-case optimal strategy for the quizmaster can be characterized by the RCAR condition (6.10). This is the condition that also characterizes worst-case optimal strategies for local loss functions and for Kelly gambling with arbitrary payoffs (by Theorem 6.10 and Lemma 6.14); the results in this chapter show that the same characterization sometimes holds for a much more general class of loss functions (as displayed in Figure 7.1). This leads to an interesting property of those (and only those) message structures, discussed in Section 7.4.3: the same strategy P^* will be optimal for the quizmaster regardless of the loss function.

Motivated by the importance and simplicity of the RCAR condition, in Section 7.5 we explore the problem of efficiently computing an RCAR strategy for the quizmaster. Depending on the messages structure, this may still be a hard problem, and we fully solve it only for a small class of message structures $\mathcal{Y}$. We encounter several other classes of messages structures in Sections 7.5.1 and 7.5.3 (also illustrated in Figure 7.1), and find some interesting properties of these classes. The topic of efficient algorithms will be explored more thoroughly in Chapter 8.

We will look at the game from the perspective of the quizmaster, and consider worst-case optimal strategies P^* for him. In games for which a Nash equilibrium exists, the contestant's worst-case optimal strategies can be found easily once we know P^* and a KT-vector certifying its optimality as in Theorem 6.3: given a KT-vector, Q^* can be constructed message-by-message to satisfy the condition in Theorem 6.7. This is even easier in the case of proper loss functions, where for each y with $P^*(y) > 0$, an optimal response is simply $Q^*_{|y} = P^*(\cdot | y)$. Another advantage of looking at the game from the quizmaster's side is that our Theorem 6.3 characterizing worst-case optimal P^* requires weaker conditions than Theorem 6.7 characterizing worst-case optimal Q^* .

The proofs of all lemmas and theorems can be found in the appendix at the end of this chapter.

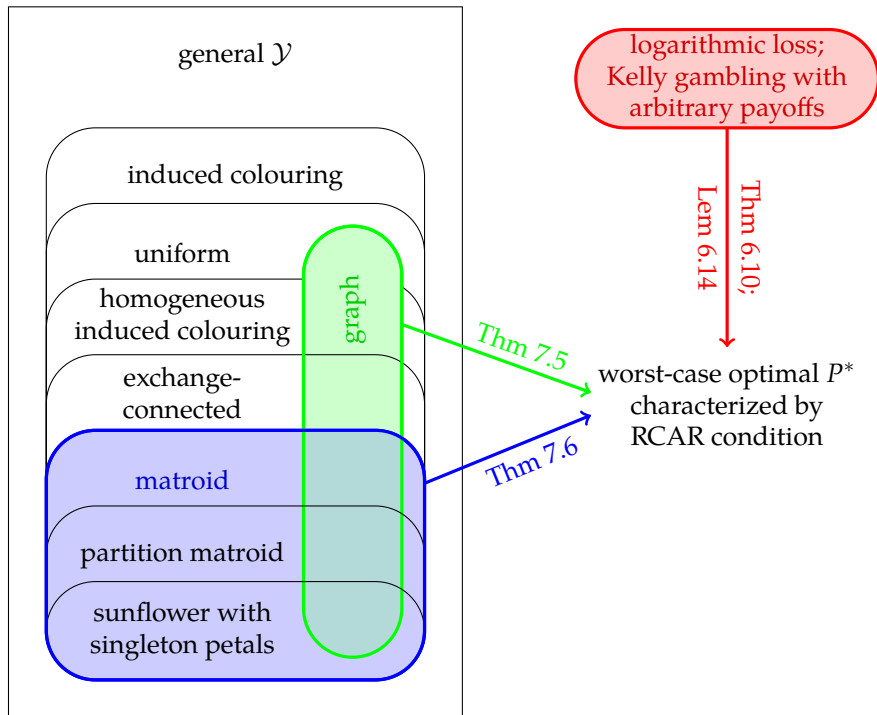


Figure 7.1: Overview of classes of message structures (compare to Table 6.1). One of the main results of this chapter is that for two classes of message structures, regardless of which loss function is used, we have the same RCAR characterization of worst-case optimal strategies for the quizmaster that we saw in Theorem 6.10 for a class of loss functions which includes logarithmic loss. Additionally, Theorem 7.7 shows that the same is not generally true for games with other message structures. The other classes shown in this figure are defined and explored in Section 7.5. (The border around ‘homogeneous induced colouring’ vanishes behind ‘graph’ because any graph with a homogeneous induced colouring is also (exchange-)connected.)

7.2 Decomposition of games

For some message structures, regardless of the marginal and loss function, the problem of finding a worst-case optimal strategy for the quizmaster can be solved by considering a smaller message structure instead. It will be useful to look at such simplifications first, so that in the rest of this chapter we will only need to deal with message structures that have already been simplified.

We have already seen one example of the type of result we are looking for earlier, in Lemma 6.2 on page 126, where we saw that if a message is *dominated* by another (meaning that it is a subset of the other), then the quizmaster always has a worst-case optimal strategy that assigns probability 0 to the dominated message.

7.2.1 Decomposition and connected games

Connectivity is a fundamental concept from graph theory. However, in general, our message structures are not graphs, but *hypergraphs*. Like an ordinary graph, a hypergraph is defined by a set of nodes and a set of edges, but the edges are allowed to be arbitrary subsets of the nodes; in a graph, all edges must contain exactly two nodes. Thus for a probability updating game, we can talk about the hypergraph $(\mathcal{X}, \mathcal{Y})$, having the outcomes as its nodes and the messages as its edges.

The terminology of connectivity can be generalized from graphs to hypergraphs (Schrijver, 2003a). We will say that a game is connected if its underlying hypergraph is connected. This leads to the following definitions.

If for some game $\mathcal{G} = (\mathcal{X}, \mathcal{Y}, p, L)$, there is a set $\emptyset \subsetneq S \subsetneq \mathcal{X}$ such that for each message y , either $y \subseteq S$ or $y \subseteq \mathcal{X} \setminus S$, then the game can be *decomposed* into two games $\mathcal{G}_1 = (\mathcal{X}_1, \mathcal{Y}_1, p^{(1)}, L)$ and $\mathcal{G}_2 = (\mathcal{X}_2, \mathcal{Y}_2, p^{(2)}, L)$ with $\mathcal{X}_1 = S$, $\mathcal{X}_2 = \mathcal{X} \setminus S$, $\mathcal{Y}_i = \{y \in \mathcal{Y} \mid y \subseteq \mathcal{X}_i\}$, and $p^{(i)}(x) = p(x) / \sum_{x' \in \mathcal{X}_i} p(x')$. If no such set S exists, we say the game $\mathcal{G}$ is *connected*.

Lemma 7.1 (Decomposition). *If a game $\mathcal{G}$ can be decomposed into $\mathcal{G}_1$ and $\mathcal{G}_2$ as described above, and its loss function L is such that H_L is finite and continuous and $H_L(P) = \inf_{Q \in \Delta_y} \sum_{x \in y} P(x)L(x, Q)$ for each $y \in \mathcal{Y}$ and each $P \in \Delta_y$, then a strategy P^* is worst-case optimal for the quizmaster in $\mathcal{G}$ if and only if there exist worst-case optimal strategies for the quizmaster P_1^* and P_2^* in $\mathcal{G}_1$ and $\mathcal{G}_2$ respectively such that*

$$P^*(x, y) = \begin{cases} P_1^*(x, y) \cdot \sum_{x' \in \mathcal{X}_1} p(x') & \text{for } x \in \mathcal{X}_1; \\ P_2^*(x, y) \cdot \sum_{x' \in \mathcal{X}_2} p(x') & \text{for } x \in \mathcal{X}_2. \end{cases}$$

(The extra condition on L is necessary to exclude some ‘very improper’ loss functions: those that reward the contestant for predicting outcomes known to have probability 0.) In particular, if the messages of $\mathcal{G}$ form a partition of $\mathcal{X}$, then $\mathcal{G}$ can be decomposed into games that each contain only one message. In a game $\mathcal{G}$ of this form, the quizmaster has only one strategy to choose from. If the loss function is proper, naive conditioning is an optimal response to this strategy, and thus worst-case optimal.

Together with Lemma 6.2, this lemma allows us to reduce any game in which we want to find a worst-case optimal strategy for the quizmaster to a set of connected games containing no dominated messages. These reduced games will not contain any messages of size one, unless one of the games consists of only that message: a message of size one is either dominated, or it forms a trivial component containing no other messages.

7.2.2 Substitution decomposition and modules

In graph theory, the concept of connected components can be generalized to *modules*. A module of a graph is a subset of its nodes such that each node outside the module is either adjacent to all or to none of the nodes in the module (Spinrad, 2003). This concept can be generalized to hypergraphs, for example as in Möhring and Radermacher (1984), by defining a module as a set $\emptyset \subsetneq \mathcal{X}' \subseteq \mathcal{X}$ such that for all $y_1, y_2 \in \mathcal{Y}$, both with $y_i \cap \mathcal{X}' \neq \emptyset$, also $(y_1 \setminus \mathcal{X}') \cup (y_2 \cap \mathcal{X}') \in \mathcal{Y}$. The sets consisting of a single outcome and the set $\mathcal{X}$ itself are always modules, and are called *trivial modules*. Any connected component of a hypergraph is also a module.

The following lemma applies in particular if $\mathcal{X}'$ is a nontrivial module, but also somewhat more generally. However, its application is restricted to logarithmic loss. (It can be extended to other local proper loss functions L with H_L finite and continuous if L is additionally symmetric on $\mathcal{X}'$, as defined in Section 7.3.1.) Thus it will not play as big a role in the rest of this text as the ordinary decomposition lemma.

Lemma 7.2 (Substitution decomposition). *Given a game $\mathcal{G} = (\mathcal{X}, \mathcal{Y}, p, L)$ with L logarithmic loss, and a set $\emptyset \subsetneq \mathcal{X}' \subsetneq \mathcal{X}$, define two new games: the ‘inner’ game $\mathcal{G}^{in} = (\mathcal{X}^{in}, \mathcal{Y}^{in}, p^{in}, L)$, with*

$$\begin{aligned}\mathcal{X}^{in} &= \mathcal{X}'; \\ \mathcal{Y}^{in} &= \{y \cap \mathcal{X}' \mid y \cap \mathcal{X}' \neq \emptyset\}; \\ p_x^{in} &= p_x / \sum_{x \in \mathcal{X}'} p_x;\end{aligned}$$

and the ‘outer’ game $\mathcal{G}^{out} = (\mathcal{X}^{out}, \mathcal{Y}^{out}, p^{out}, L)$, with

$$\begin{aligned}\mathcal{X}^{out} &= \mathcal{X} \setminus \mathcal{X}' \cup \{x'\}; \\ \mathcal{Y}^{out} &= \{y \in \mathcal{Y} \mid y \cap \mathcal{X}' = \emptyset\} \cup \{y \setminus \mathcal{X}' \cup \{x'\} \mid y \in \mathcal{Y}, y \cap \mathcal{X}' \neq \emptyset\}; \\ p_x^{out} &= \begin{cases} p_x & \text{for } x \neq x'; \\ \sum_{x \in \mathcal{X}'} p_x & \text{for } x = x', \end{cases}\end{aligned}$$

where $x' \in \mathcal{X}'$ is an arbitrary outcome. Let p^{in}, p^{out} be worst-case optimal strategies for these games, with respective RCAR vectors q^{in}, q^{out} . If for all $y^{in} \in \mathcal{Y}^{in}, y^{out} \in \mathcal{Y}^{out}$ with $p^{in}(y^{in}) > 0, y^{out} \ni x'$ and $p^{out}(y^{out}) > 0$ in the new games, we have $y^{out} \setminus \{x'\} \cup y^{in} \in \mathcal{Y}$ in the original game, then a worst-case optimal strategy for the

original game is given by

$$P^*(y) = \begin{cases} P^{out}(y) & \text{for } y \cap \mathcal{X}' = \emptyset; \\ P^{out}(y \setminus \mathcal{X}' \cup \{x'\}) \cdot P^{in}(y \cap \mathcal{X}') & \text{otherwise;} \end{cases}$$

and RCAR vector

$$q_x = \begin{cases} q_x^{out} & \text{for } x \notin \mathcal{X}'; \\ q_{x'}^{out} \cdot q_x^{in} & \text{for } x \in \mathcal{X}'. \end{cases}$$

An example of a nontrivial module appeared in the game in Example 6.D on page 129, which has message structure $\mathcal{Y} = \{\{x_1, x_2\}, \{x_2, x_3, x_4\}\}$; there $\mathcal{X}' = \{x_3, x_4\}$ is a module. (Another nontrivial module is $\{x_1, x_3, x_4\}$.) The set $\{x_3, x_4\}$ is the simplest kind of module: there is only one message that intersects it, and so the condition is satisfied trivially. Somewhat more generally, if in a message structure, all messages y that intersect with $\mathcal{X}'$ contain all of $\mathcal{X}'$, then $\mathcal{X}'$ is a module. Applying the lemma to this case tells us that for logarithmic loss, the game can be simplified by *merging the outcomes* in $\mathcal{X}'$ into a single outcome; the original game's worst-case optimal strategy will then distribute the mass on this single outcome among the outcomes in $\mathcal{X}$ proportionally to their marginals. We also saw in Example 6.D that for loss functions other than logarithmic loss, the strategy found this way may not be worst-case optimal.

7.3 Outcome symmetry

Sometimes, the problem of finding a worst-case optimal strategy is simplified because certain 'symmetry' properties of the message structure and loss function allow us to conclude that worst-case optimal strategies satisfying an additional condition must have the same symmetries.

7.3.1 Symmetry of loss functions

We now briefly return to the topic of loss functions to define a property we will need next.

For a probability distribution $Q \in \Delta_{\mathcal{X}}$ and x_1, x_2 distinct elements of $\mathcal{X}$, define $Q^{x_1 \leftrightarrow x_2}$ as

$$Q^{x_1 \leftrightarrow x_2}(x) = \begin{cases} Q(x_2) & \text{for } x = x_1; \\ Q(x_1) & \text{for } x = x_2; \\ Q(x) & \text{otherwise,} \end{cases}$$

and similarly for a contestant's strategy $Q \in \mathcal{Q}$ by applying this transformation to the conditional for each y . We say L is *symmetric between x_1 and x_2* if for all $Q \in \Delta_{\mathcal{X}}$, we have $L(x_1, Q) = L(x_2, Q^{x_1 \leftrightarrow x_2})$ and $L(x, Q) = L(x, Q^{x_1 \leftrightarrow x_2})$ for all $x \in \mathcal{X} \setminus \{x_1, x_2\}$. If L is symmetric between x_1 and x_2 and between x_2 and x_3 , then it is also symmetric between x_1 and x_3 , because $((Q^{x_1 \leftrightarrow x_2})^{x_2 \leftrightarrow x_3})^{x_1 \leftrightarrow x_2} = Q^{x_1 \leftrightarrow x_3}$. In words: we can apply the first symmetry, then the second, then the

first again to find that we have exchanged x_1 and x_3 . We also consider any loss function to be symmetric between x and x for any x . So this symmetry of L is an equivalence relation on $\mathcal{X}$, and we are justified in talking about L being symmetric on sets $S \subseteq \mathcal{X}$, meaning that all pairs of elements of that set can be exchanged. If L is symmetric on $\mathcal{X}$, we say it is *fully symmetric*.

The loss functions we have seen so far were fully symmetric with the exception of the loss function in Example 6.K. The affine transformations of loss functions discussed at the end of Section 6.5.2 may change the symmetries of a loss function, while they do not change which strategies are worst-case optimal for the two players. This means that sometimes, an asymmetric loss functions can be transformed into an essentially equivalent loss function with better symmetry properties. The loss function from Example 6.K cannot be transformed this way. Other loss functions that may exhibit this kind of *inherent* asymmetry are given in the following two examples.

Example 7.A (Matrix loss). Given a $[0, \infty)$ -valued $\mathcal{X} \times \mathcal{X}$ matrix of losses A , define *hard matrix loss* by

$$L(x, Q) = \begin{cases} A_{x, x'} & \text{if } Q(x') = 1 \text{ for some } x'; \\ \infty & \text{otherwise.} \end{cases}$$

This generalizes hard 0-1 loss, which is obtained for the matrix A with zeroes on the diagonal and ones elsewhere (except that the definition above may give infinite loss for some Q , but a rational contestant would never use such Q). It is symmetric between x_1 and x_2 if and only if swapping row x_1 with x_2 and column x_1 with x_2 results in matrix A again; that is, if and only if $A_{x_1, x_1} = A_{x_2, x_2}$, $A_{x_1, x_2} = A_{x_2, x_1}$, $A_{x', x_1} = A_{x', x_2}$, and $A_{x_1, x'} = A_{x_2, x'}$, for all $x' \in \mathcal{X} \setminus \{x_1, x_2\}$.

We can also define randomized matrix loss as an analogous generalization of randomized 0-1 loss, by taking an expectation over Q in hard matrix loss:

$$L(x, Q) = \sum_{x' \in \mathcal{X}} Q(x') A_{x, x'}.$$

It has the same symmetry properties as hard matrix loss. The proof of Proposition 6.6 also applies to randomized matrix loss without modification, showing that a Nash equilibrium exists in games using this loss function.

Example 7.B (Skewed logarithmic loss). Fix a vector $c \in \mathbf{R}_{\geq 0}^{\mathcal{X}}$, and define the function $F : \Delta_{\mathcal{X}} \rightarrow \mathbf{R}_{\geq 0}$ by

$$F(P) := - \sum_{x \in \mathcal{X}} c_x P(x) \log P(x).$$

This is a sum of differentiable concave functions, and therefore differentiable and concave; if $c \in \mathbf{R}_{> 0}^{\mathcal{X}}$, it is strictly concave (in fact, it is also strictly concave if c contains a single 0). We use the construction of *Bregman scores* in Grünwald

and Dawid (2004, Section 3.5.4) to construct a proper loss function L having F as its generalized entropy, and find

$$L(x, Q) = F(Q) + (e_x - Q) \cdot \nabla F(Q) = -c_x(1 + \log Q(x)) + \sum_{x' \in \mathcal{X}} c_{x'} Q(x'),$$

where e_x is the distribution that puts all mass on x . This loss function is strictly proper if H_L is strictly concave. Unlike logarithmic loss and its affine transformations, it is not local for $|\mathcal{X}| > 2$. Also, it is not generally fully symmetric, but is symmetric between pairs of outcomes $x_1, x_2 \in \mathcal{X}$ with $c_{x_1} = c_{x_2}$.

7.3.2 Symmetry of KT-vectors

Using the definition of symmetry of loss functions introduced in the previous section, we can now state the following lemma.

Lemma 7.3 (Loss exchange). *Consider a game with $y_1, y_2 \in \mathcal{Y}$, $y_1 \setminus y_2 = \{x_1\}$, $y_2 \setminus y_1 = \{x_2\}$, H_L finite and continuous and L symmetric between x_1 and x_2 . If a worst-case optimal strategy P^* for the quizmaster exists with $P^*(x_1, y_1) > 0$, then all KT-vectors λ^* satisfy $\lambda_{x_1}^* \leq \lambda_{x_2}^*$.*

When two messages $y_1, y_2 \in \mathcal{Y}$ satisfy $y_1 \setminus y_2 = \{x_1\}$ and $y_2 \setminus y_1 = \{x_2\}$, we say that they differ by the *exchange* of one outcome.

In order to find worst-case optimal strategies, we would like to be able to relate $\lambda_{x_1}^*$ to $\lambda_{x_2}^*$ whenever $P^*(y_1) > 0$, but the previous lemma requires something stronger: that $P^*(x_1, y_1) > 0$. We call a strategy P *degenerate* if there exist $y_1, y_2 \in \mathcal{Y}$, $y_1 \setminus y_2 = \{x_1\}$, $y_2 \setminus y_1 = \{x_2\}$ as in the above lemma with $P(y_1) > 0$ and $P(y_2) > 0$ but $P(x_1, y_1) = 0$. Otherwise, P is called *nondegenerate*; then $P(y_1) > 0$, $P(y_2) > 0$ implies $P(x_1, y_1) > 0$, $P(x_2, y_2) > 0$.

We similarly want a term for the symmetry conditions on L that allow us to apply Lemma 7.3 to any pair of messages in some set $\mathcal{Y}' \subseteq \mathcal{Y}$ satisfying the statement of the lemma. We say L is *symmetric with respect to exchanges in $\mathcal{Y}'$* if L is symmetric between any pair of outcomes x_1, x_2 such that messages $y_1, y_2 \in \mathcal{Y}'$ exist with $y_1 \setminus y_2 = \{x_1\}$ and $y_2 \setminus y_1 = \{x_2\}$.

Lemma 7.4 (Transfer of λ^*). *Consider a game and a worst-case optimal strategy P^* for the quizmaster such that H_L finite and continuous and L symmetric with respect to exchanges in $\{y \in \mathcal{Y} \mid P^*(y) > 0\}$. Then we have one of the following:*

- If P^* is nondegenerate, then $\lambda_{x_1}^* = \lambda_{x_2}^*$ for all x_1, x_2 such that messages $y_1, y_2 \in \mathcal{Y}$ exist with $y_1 \setminus y_2 = \{x_1\}$, $y_2 \setminus y_1 = \{x_2\}$, $P^*(y_1) > 0$, and $P^*(y_2) > 0$;
- If P^* is degenerate, then a nondegenerate worst-case optimal strategy P' exists with $\{y \in \mathcal{Y} \mid P'(y) > 0\} \subsetneq \{y \in \mathcal{Y} \mid P^*(y) > 0\}$.

Example 7.C (Degenerate P^*). Consider the game

P^*	x_1	x_2	x_3	x_4
y_1	1/8	2/8	—	0
y_2	—	2/8	1/8	0
y_3	0	—	0	2/8
p_x	1/8	4/8	1/8	2/8

with loss function

$$L(x, Q) = \begin{cases} 1 - Q(x) & \text{for } x \neq x_4; \\ 1 & \text{for } x = x_4. \end{cases}$$

This instance of randomized matrix loss (introduced in Example 7.A; here, the loss matrix A has $A_{x,x} = 0$ for $x \in \{x_1, x_2, x_3\}$ and equals 1 elsewhere) gives the contestant no incentive to predict $Q(x_4) > 0$, always assigning him the same loss if that outcome does occur. It is not fully symmetric, but it is symmetric between x_1, x_2 and x_3 , so symmetric with respect to exchanges in $\mathcal{Y}$. The strategy P^* given in the table is worst-case optimal, as witnessed by $\lambda^* = (1, 0, 1, 1)$. Though y_1 and y_3 differ by the exchange of one outcome (x_2 for x_3) and have positive probability, we have $\lambda_2^* \neq \lambda_3^*$; similar for y_2 and y_3 and $\lambda_1^* \neq \lambda_2^*$.

In this example, the quizmaster has worst-case optimal strategies $P' \neq P^*$ with $P'(y_3) = 0$, dividing the mass $P^*(x_4, y_3)$ among $P'(x_4, y_1)$ and $P'(x_4, y_2)$.

7.4 The RCAR characterization for general loss functions

We saw in Theorem 6.10 that for logarithmic loss, worst-case optimal strategies for the quizmaster can be characterized in terms of a simple condition, the *RCAR condition* (6.10). We also saw that sometimes (in Examples 6.B and 6.C on pages 127 and 129, but not in Example 6.D), those same strategies were also worst-case optimal for other loss functions. This suggests that even for some types of games where Theorem 6.10 does not apply, it is possible to recognize worst-case optimal strategies using the easily verifiable RCAR condition. We show that there are two classes of message structures in which this is possible regardless of the marginal, and explore the consequences in Section 7.4.3.

7.4.1 Graph games

The first of these classes consist of all message structures $\mathcal{Y}$ for which each message contains at most two outcomes. After removing singleton messages (which are either dominated or are decomposable from the rest of the game), we have $|y| = 2$ for all $y \in \mathcal{Y}$. This corresponds to a simple undirected graph (that is, a graph containing no loops or multiple edges) with a node for each outcome in $\mathcal{X}$ and an edge for each message in $\mathcal{Y}$. For this reason, a game where each message in $\mathcal{Y}$ contains at most two outcomes is called a *graph game*.

Many games we saw in the examples in Chapter 6 were graph games. Their underlying graphs are shown in Figure 7.2.

Theorem 7.5 (RCAR for graph games). *If each message in $\mathcal{Y}$ contains at most two outcomes and $P^* \in \mathcal{P}$ is an RCAR strategy, then P^* is worst-case optimal for all L symmetric with respect to exchanges in $\{y \in \mathcal{Y} \mid |y| = 2\}$ with H_L finite and continuous. If additionally H_L is strictly concave, only such P^* are worst-case optimal for L .*

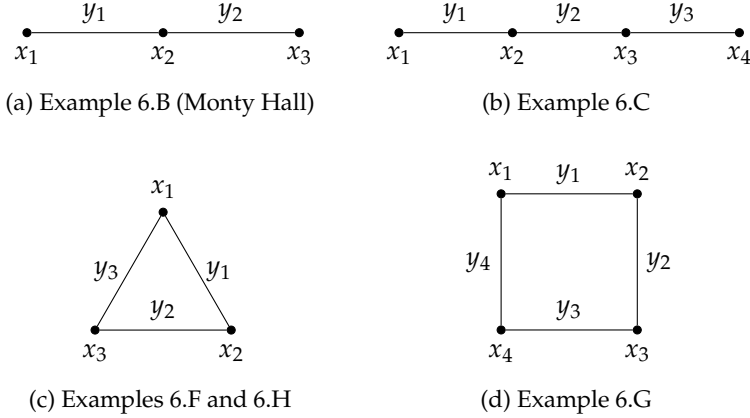


Figure 7.2: Underlying graphs of the graph games seen in Chapter 6

The statement of the theorem is very similar to that of Theorem 6.10 in Section 6.5.2, and the restrictions on L in the present theorem (except for symmetry) were also seen in the previous theorem. Sufficient conditions for these restrictions to hold were given by Lemma 6.1 in Section 6.3 (H_L finite and continuous) and Lemma 6.13 in Section 6.5.2 (strict concavity of H_L).

The intuition behind the proof is that for binary predictions Q , the probability assigned by Q to one outcome determines the probability Q assigns to the other outcome. Thus all loss functions are essentially local when used to assess such predictions, and their behaviour is similar to logarithmic loss.

7.4.2 Matroid games

The other class is that of *matroid games*. A *matroid* over a finite ground set $\mathcal{X}$ can be defined by a nonempty family $\mathcal{Y}$ of subsets of $\mathcal{X}$ (the *bases* of the matroid) satisfying the *basis exchange* property (Oxley, 2011, Corollary 1.2.5): for all $y_1, y_2 \in \mathcal{Y}$ and $x_1 \in y_1 \setminus y_2$,

$$(y_1 \setminus \{x_1\}) \cup \{x_2\} \in \mathcal{Y} \text{ for some } x_2 \in y_2 \setminus y_1. \quad (7.1)$$

In words, for any pair of messages, if an outcome that is not in the second message is removed from the first message, it must be possible to replace it by an outcome from the second message that is not in the first message, in such a way that the resulting set of outcomes is again a message.

A matroid game is a game in which $\mathcal{Y}$ is the set of bases of a matroid. The Monty Hall game (Example 6.B) is a matroid game: taking one of the two messages and replacing the outcome unique to it by the only other outcome will result in the other message. By our definition of a game, it is required in addition to (7.1) that each element of the ground set $\mathcal{X}$ of the matroid occurs in some basis.

Many alternative characterizations of matroids exist. For example, a matroid with ground set $\mathcal{X}$ and bases $\mathcal{Y}$ can also be represented by its family of *independent sets* $\mathcal{I} = \{I \subseteq \mathcal{X} \mid I \subseteq y \text{ for some } y \in \mathcal{Y}\}$, and a different set of axioms analogous to (7.1) characterizes whether a given set $\mathcal{I}$ is the family of independent sets of some matroid.

The concept of a matroid was introduced by Whitney (1935) to study the abstract properties of the notion of dependence, as seen for example in linear algebra and graph theory (explained below). Different characterizations of the concept, applied to different examples, were given independently by other authors, but then turned out to be equivalent to matroids. One field where matroids play an important role is combinatorial optimization. We refer to Schrijver (2003b, Section 39.10b) for extensive historical notes.

We give two example classes of matroids, taken from Schrijver (2003b, Section 39.4):

- Given an $m \times n$ matrix A over some vector space, let $\mathcal{X} = \{1, 2, \dots, n\}$ and $\mathcal{I}$ the family of all subsets I of $\mathcal{X}$ such that the set of column vectors with index in I is linearly independent. Then $\mathcal{I}$ is the family of independent sets of a matroid. A subset that spans the column space of A is a basis of this matroid.
- Given a simple undirected graph G , let $\mathcal{X}$ be its set of edges and $\mathcal{I}$ consist of all acyclic subsets of $\mathcal{X}$. Then $\mathcal{I}$ is the family of independent sets of a matroid. This matroid is called the *cycle matroid* of G . The bases are the maximal independent sets; if G is connected, these are its spanning trees.

One interesting class of games for which $\mathcal{Y}$ are the bases of a matroid is the class of *negation games*. In such a game, each element of $\mathcal{Y}$ is of the form $\mathcal{X} \setminus \{x\}$ for some x . (Not all sets of this form need to be in $\mathcal{Y}$.) Thus the quizmaster will tell the contestant, “The true outcome is *not* x ,” as in the original Monty Hall problem where one of the three doors is opened to reveal a goat. A family $\mathcal{Y}$ of this form satisfies (7.1) trivially: for y_1, y_2 distinct elements of $\mathcal{Y}$, there is only one choice for each of x_1 and x_2 , and with these choices we get $(y_1 \setminus \{x_1\}) \cup \{x_2\} = y_2 \in \mathcal{Y}$.

Another class of matroids is formed by the *uniform matroids*, in which *every* set of some fixed size k is a basis. These also have a natural interpretation when they occur as the message structure of a game: the quizmaster is allowed to leave any set of k doors shut.

As the following theorem shows, matroid games share with graph games the property that RCAR strategies are worst-case optimal for a wide variety of loss functions. Section 7.5 will provide an intuition of why these message structures have this property, and the proof of the theorem uses some terminology introduced in that section.

Theorem 7.6 (RCAR for matroid games). *If $\mathcal{Y}$ are the bases of a matroid and $P^* \in \mathcal{P}$ is an RCAR strategy, then P^* is worst-case optimal for all L symmetric with respect to exchanges in $\mathcal{Y}$ with H_L finite and continuous. If additionally H_L is strictly concave, only such P^* are worst-case optimal for L .*

7.4.3 Loss invariance

We saw in the preceding sections that in graph and matroid games, worst-case optimal strategies for the quizmaster are characterized by the RCAR property. This property does not depend on what loss function is used in the game (though the theorems do put some conditions on the loss function, such as some symmetry requirements). Consequently, in such games, strategies exist that are worst-case optimal regardless of what loss function is used (at least, for a large class of loss functions). We call this phenomenon *loss invariance*.

For such message structures, we can really think of the worst-case optimal strategies as ‘conditioning’ (as a purely probability-based operation) rather than as worst-case optimal strategies for some game. This conditioning operation can be seen as the generalization of naive conditioning to message structures other than partitions (where naive conditioning gives the right answer). Unlike naive conditioning, which requires just the distribution p and the message y to compute $P(x \mid y)$, we also need the message structure $\mathcal{Y}$ to compute that conditional probability. But like naive conditioning, we do not need to fix a loss function in order to talk about the worst-case optimal prediction of x given a message y .

A subtlety appears when improper loss functions are considered. Our theorems show that the worst-case optimal strategies for the quizmaster are characterized independently of the loss function; however, the worst-case optimal strategies for the contestant will not necessarily coincide with these if the loss function is not proper. In this case, loss invariance tells us that the loss function does not affect *what the contestant should believe* about the true outcome, but it may affect how the contestant translates this belief into a prediction.

In the cases of graph and matroid games, our analysis of worst-case optimal strategies becomes more widely applicable in situations where the probability updating game is really played by two players (as opposed to being a theoretical tool for defining safe updating strategies):

- the same strategies continue to be worst-case optimal if the two players use different loss functions (so that the game is no longer zero-sum);
- both players will be able to play optimally without knowing the loss function(s) in use.

This is true for the Monty Hall game (Example 6.B), which lies in the intersection of graph and matroid games. This provides some justification for the prevailing intuition that the Monty Hall problem should be analysed using probability theory, without mention of loss functions.

Theorems 7.5 and 7.6 apply only to loss functions that are sufficiently symmetric and for which H_L is continuous and finite. We make no claim about the question whether RCAR strategies are also worst-case optimal for loss functions that do not satisfy these properties. However, note that by Lemma 6.14, sometimes affine transformations can be used to convert an asymmetric loss function into a symmetric one without affecting the players’ strategies.

Lemma 6.14 also shows that a limited form of loss invariance holds regardless of the message structure. If the players are using different affine transformations of the same loss function (for example, of logarithmic loss; this corresponds to Kelly gambling where the pay-offs for the contestant are different from those for the quizmaster), both players can play optimally without knowing the transformations in use.

An obvious question that remains is: are there any other classes of message structure for which we have loss invariance? This is answered in the negative by the following theorem.

Theorem 7.7. *If a connected game containing no dominated messages is neither a matroid game nor a graph game, then there exists a marginal such that no strategy P for the quizmaster is worst-case optimal for both logarithmic loss and Brier loss.*

7.5 Finding RCAR strategies

We have now seen three situations in which worst-case optimal strategies for the quizmaster can be characterized using the RCAR condition: if L is local and proper (such as logarithmic loss; see Theorem 6.10), if $\mathcal{Y}$ is a graph (Theorem 7.5), and if $\mathcal{Y}$ is a matroid (Theorem 7.6). Thus in order to find a worst-case optimal strategy, it would be helpful to be able to find RCAR strategies.

We also saw in Theorem 7.7 that if $\mathcal{Y}$ is neither a graph nor a matroid, there exist pairs of marginals and loss functions for which RCAR strategies are not worst-case optimal. But this does not hold for all marginals and loss functions, so even if L is not logarithmic loss and $\mathcal{Y}$ is not a graph or a matroid (i.e. if no RCAR theorem holds), it may be worthwhile to look for an RCAR strategy, and check if it is optimal using the results of Chapter 6.

In this section, we establish a computational procedure that tries to find an RCAR strategy given a message structure and a marginal. We do not call this procedure an ‘algorithm’ because, unless the input satisfies special conditions, it may not be applicable, or give an inconclusive answer. In Chapter 8, we will see algorithms that efficiently find RCAR strategies. However, these algorithms are restricted to the cases of graph and matroid games, and understanding those algorithms does not give an understanding of many of the problems we may run into when looking for RCAR strategies. To gain such understanding, the present section is more useful: in the course of developing our computational procedure, we acquire more insight into why sometimes the RCAR property characterizes worst-case optimal strategies P^* , and what makes graph and matroid games special.

7.5.1 Induced colourings

Fix a set $\mathcal{Y}' \subseteq \mathcal{Y}$ with $\bigcup_{y \in \mathcal{Y}'} y = \mathcal{X}$, and assume that an RCAR strategy P exists with support $\mathcal{Y}_P := \{y \in \mathcal{Y} \mid P(y) > 0\}$ equal to $\mathcal{Y}'$. (For example, we may in many cases take $\mathcal{Y}' = \mathcal{Y}$.) It follows from the RCAR property that $P(x \mid$

$y) > 0$ for all $x \in y \in \mathcal{Y}_P$, so in particular that P is nondegenerate. We will now consider different properties of $\mathcal{Y}_P$ that may help us find P . The classes of message structures defined by these properties, and the inclusion relations between them that we establish here, were shown graphically in Figure 7.1 on page 151, and examples are given in Figure 7.3.

Consider the system of linear equations

$$\sum_{x \in y} q_x = 1 \text{ for all } y \in \mathcal{Y}'. \quad (7.2)$$

A positive solution q of this system shows the existence of an RCAR strategy with support $\mathcal{Y}'$ and RCAR vector q for some marginal p ; the nonexistence of such a solution implies that no such strategies exist for any positive marginal. (A similar system is studied in the CAR literature, where it plays a role in characterizing message structures that admit a CAR coarsening mechanism; see Grünwald and Halpern (2003); Jaeger (2005b); Gill and Grünwald (2008). Since we study RCAR rather than CAR, the roles of outcomes and messages are reversed here. We say more about this correspondence in Section 7.6.1, where we relate some of the classes defined below to classes defined in the CAR literature.)

Define a *colouring* as a partition of $\mathcal{X}$. We say a colouring is *induced* by a set of messages $\mathcal{Y}'$ if the system of linear equations (7.2) has at least one solution q with $q_x > 0$ for all x , and x, x' are in the same class of the colouring ('have the same colour') if and only if $q_x = q_{x'}$ for all such solutions to that system. If the system has at least one positive solution, then the colouring induced by $\mathcal{Y}'$ is unique; otherwise, there is no induced colouring.

We say a colouring is *homogeneous on $\mathcal{Y}'$* if the number of outcomes of each colour is the same for every message in $\mathcal{Y}'$ (for example, if each message consists of one 'red' and two 'blue' outcomes). This is only possible if $\mathcal{Y}'$ is *uniform*: all messages in $\mathcal{Y}'$ have the same size. We are interested in $\mathcal{Y}'$ whose induced colouring is homogeneous. One class of such $\mathcal{Y}'$ that is easy to recognize consists of those $\mathcal{Y}'$ that are *exchange-connected*: for each pair of messages in $\mathcal{Y}'$, there is a path of messages in $\mathcal{Y}'$ (an *exchange-path*) whose adjacent messages differ by the exchange of one outcome as in the conditions of Lemma 7.3 and 7.4.

Figure 7.3 illustrates these definitions with a few examples. The tables are of the same form as those used to display message structures in previous examples, except that the cells now show a colouring instead of a strategy for the quizmaster.

The message structure shown in Figure 7.3a has no induced colouring: any solution of (7.2) must have $q_{x_2} = 0$, so there is no positive solution, and it follows that no RCAR strategy P exists with $P(y) > 0$ for all $y \in \mathcal{Y}'$. On the other hand, any uniform game has an induced colouring, because there is at least one solution to (7.2):

$$q_x = 1/k \quad \text{for all } x \in \mathcal{X}, \quad (7.3)$$

where k is the size of the game's messages.

	x_1	x_2	x_3	x_4	x_5
y_1	*	*	*	—	—
y_2	—	—	*	*	—
y_3	—	—	—	*	*
y_4	*	—	—	—	*

(a) No induced colouring

	x_1	x_2	x_3	x_4
y_1	*	*	—	—
y_2	—	*	*	*

(b) Induced colouring but not uniform

	x_1	x_2	x_3	x_4	x_5	x_6
y_1	*	*	*	—	—	—
y_2	—	—	*	*	*	—
y_3	*	—	—	—	*	*

(c) Uniform but induced colouring not homogeneous

	x_1	x_2	x_3	x_4	x_5	x_6
y_1	*	*	*	—	—	—
y_2	—	—	*	*	*	—
y_3	*	—	—	—	*	*
y_4	—	*	—	*	—	*

(d) Homogeneous induced colouring but not exchange-connected

	x_1	x_2	x_3	x_4	x_5
y_1	*	*	*	—	—
y_2	—	*	*	*	—
y_3	—	—	*	*	*

(e) Exchange-connected but not a matroid

	x_1	x_2	x_3	x_4	x_5
y_1	*	*	*	—	—
y_2	*	*	—	*	—
y_3	*	—	*	*	—
y_4	*	—	*	—	*
y_5	*	—	—	*	*

(f) Matroid

Figure 7.3: Examples of messages structures and their induced colourings

Figures 7.3b and 7.3c are examples of message structures that do have an induced colouring, but one that is not homogeneous. In both these examples, all outcomes have different colours in the induced colouring, because no pair of outcomes necessarily has the same value of q in a solution of (7.2). The message structure shown in Figure 7.3c will be revisited in Example 7.D in the next section.

The three remaining message structures do have homogeneous induced colourings. Figure 7.3d shows that it is possible for a message structure to have a homogeneous induced colouring without being exchange-connected. In this message structure, which adds the message y_4 to the structure in Figure 7.3c, it is still the case that each pair of messages differs by *two* exchanges. Yet the added message changes the induced colouring: for example, $q_{x_1} = q_{x_4}$ follows because by the equalities from (7.2) on y_1 and y_3 , $1 - q_{x_1} = q_{x_2} + q_{x_3} = q_{x_5} + q_{x_6}$, and by y_2 and y_4 , $1 - q_{x_4} = q_{x_3} + q_{x_5} = q_{x_2} + q_{x_6}$; thus $2 - 2q_{x_1} = 2 - 2q_{x_4} = q_{x_2} + q_{x_3} + q_{x_5} + q_{x_6}$.

The message structure shown in Figure 7.3e is exchange-connected. For such structures, it easy to determine the induced (homogeneous) colouring: if messages y_1, y_2 differ by the exchange of one outcome (x_1 for x_2), then any

solution of (7.2) must satisfy $q_{x_1} = q_{x_2}$, so such x_1, x_2 must be the same colour. Any vector q that satisfies all these equalities and satisfies $\sum_{x \in y} q_x = 1$ for any one message $y \in \mathcal{Y}'$ satisfies (7.2) for all messages in $\mathcal{Y}'$, so this determines the induced colouring. This colouring is clearly homogeneous on any pair of message that differ by the exchange of one outcome; because exchange-paths exist between all pairs of messages, it follows that the induced colouring of an exchange-connected game is homogeneous.

The structure in Figure 7.3e is not a matroid: there is no outcome in $y_3 \setminus y_1$ that can be added to $y_1 \setminus \{x_2\} = \{x_1, x_3\}$ to make a message. If a message $y_4 = \{x_1, x_3, x_5\}$ is added, the resulting message structure is a matroid.

Finally, the class of matroid games is a subclass of exchange-connected games: (7.1) requires the existence of not just one, but possibly many different exchange-paths between any pair of messages. Negation matroids and uniform matroids were already described in the previous section as examples of matroids; Figure 7.3f shows another example.

The following lemma gives two alternate characterizations of the induced colouring of a matroid. The first of these is in terms of a concept from matroid theory: the colour classes of the induced colouring coincide with the *2-connected components* of the matroid. (We refer to Oxley (2011) for the definition. In matroid theory, these components are usually simply called ‘connected components’, but we keep the 2 to avoid confusion with the notion of connectedness used in Lemma 7.1.) We observed above (when discussing Figure 7.3e) that if messages exist that differ in the exchange of one outcome, then the outcomes being exchanged must be the same colour. The second characterization shows that for matroids, the converse also holds. Finally, the lemma shows that every colour class of a matroid is a module as defined in Section 7.2.2 (though not every module is a colour class).

Lemma 7.8 (Matroid colouring). *Given a matroid $(\mathcal{X}, \mathcal{Y})$ and two elements $x_1, x_2 \in \mathcal{X}$, the following statements are equivalent:*

1. x_1 and x_2 are in the same colour class of the induced colouring of $\mathcal{Y}$;
2. x_1 and x_2 are in the same 2-connected component of $(\mathcal{X}, \mathcal{Y})$;
3. There exist $y_1, y_2 \in \mathcal{Y}$ such that $y_1 \setminus y_2 = \{x_1\}$ and $y_2 \setminus y_1 = \{x_2\}$.

Further, if $C \subseteq \mathcal{X}$ is a colour class of the induced colouring of $\mathcal{Y}$, then C is a module.

7.5.2 A computational procedure

Consider the case that $\mathcal{Y}'$ induces a homogeneous colouring, and assume as before that an RCAR strategy P exists with $\mathcal{Y}_P = \mathcal{Y}'$. Then the RCAR vector q must be a solution of the linear system (7.2). Additionally, P (and thus q) must agree with the game’s marginal p . These constraints allow us to compute the vector q directly.

Let S be the set of all outcomes with a particular colour. Then there is some value q_S such that $P(x | y) = q_x = q_S$ for all $x \in S, x \in y \in \mathcal{Y}$. Let $k_S = |S \cap y|$ (this is independent of y by homogeneity). We must have

$$k_S q_S = k_S \sum_y P(y) q_S = \sum_{x \in S} \sum_{y \ni x} P(y) P(x | y) = \sum_{x \in S} p_x,$$

so that q_S can be computed by

$$q_S = \frac{1}{k_S} \sum_{x \in S} p_x. \quad (7.4)$$

A simple case is when the induced colouring assigns the same colour to all outcomes: then we see that as in (7.3), we get $q_x = 1/k$ for all $x \in \mathcal{X}$, where k is the size of the messages. When a colour consists of just one outcome x (which must then be an element of every message for the colouring to be homogeneous), we find $q_x = p_x$.

If an RCAR strategy P exists with $\mathcal{Y}_P = \mathcal{Y}'$ where $\mathcal{Y}'$ induces a homogeneous colouring, then P must have the vector q given by (7.4) as its RCAR vector. However, it may be the case that no such strategy exists. To find P if it exists, we still need to determine the $P(y)$'s. We can find a nonnegative solution or determine that no nonnegative solution exists by solving the following linear programming problem (which we can do in polynomial time):

$$\begin{aligned} & \text{maximize} && \sum_{y \in \mathcal{Y}} r_y \\ & \text{subject to} && \sum_{y \ni x} r_y \leq \frac{p_x}{q_x} \quad \text{for all } x \in \mathcal{X}, \end{aligned} \quad (7.5)$$

with $r \in \mathbf{R}_{\geq 0}^{\mathcal{Y}}$. If a vector achieving $\sum_{y \in \mathcal{Y}} r_y = 1$ is found, we have a strategy P with r as the marginal on messages ($P(x, y) = q_x r_y$ for all $x \in y$). If no vector r achieves the value 1, there is no RCAR strategy P satisfying the assumption $\mathcal{Y}_P = \mathcal{Y}'$.

Now we may want to apply this procedure in practice to find an RCAR strategy for a given game. (Note that by Lemma 6.11, such a strategy always exists.)

When doing so we encounter two problems: we need to provide the procedure with an $\mathcal{Y}'$ such that $\bigcup \mathcal{Y}' = \mathcal{X}$, and even if we have an idea about what $\mathcal{Y}'$ to take, it may not have a homogeneous induced colouring. Still, let us investigate what happens if we just guess an $\mathcal{Y}'$. We will then encounter one of the cases 1, 2a-2c which we now describe. Briefly, in case 1, the procedure cannot be used because q cannot be determined, and in case 2a and 2b it gives an inconclusive result; in case 2c we have success. We now consider each case in detail.

1. $\mathcal{Y}'$ has no homogeneous induced colouring.

In this case, the procedure is not applicable. Indeed, finding an RCAR vector may be a more difficult type of problem, as illustrated by the following example

which uses the message structure from Figure 7.3c. (This example is a uniform game; the class of uniform games is the smallest class among those identified in the previous section that contains the class of games with a homogeneous induced colouring.)

Example 7.D (Irrational RCAR vector). Consider...

P	x_1	x_2	x_3	x_4	x_5	x_6
y_1	1/10	1/10	$\frac{3}{10} - \frac{1}{10}\sqrt{5}$	—	—	—
y_2	—	—	$\frac{1}{10}\sqrt{5} - \frac{1}{10}$	1/5	$\frac{1}{10}\sqrt{5} - \frac{1}{10}$	—
y_3	1/10	—	—	—	$\frac{3}{10} - \frac{1}{10}\sqrt{5}$	1/10
q_x	$\frac{1}{4} + \frac{1}{20}\sqrt{5}$	$\frac{1}{4} + \frac{1}{20}\sqrt{5}$	$\frac{1}{2} - \frac{1}{10}\sqrt{5}$	$\frac{1}{5}\sqrt{5}$	$\frac{1}{2} - \frac{1}{10}\sqrt{5}$	$\frac{1}{4} + \frac{1}{20}\sqrt{5}$
p_x	1/5	1/10	1/5	1/5	1/5	1/10

...with marginal on the messages $P(y_1) = P(y_3) = \frac{1}{2} - \frac{1}{10}\sqrt{5}$, $P(y_2) = \frac{1}{5}\sqrt{5}$. The vector q is also shown, so the RCAR property can be easily verified. We see that the RCAR strategy P and RCAR vector q (both of which are unique) contain irrational numbers, while the marginal p was rational. The solution techniques used in this section (the formula (7.4) for q and linear optimization for (7.5)) do not yield irrational results when given rational inputs, so this example shows that these techniques will not suffice in general for games that do not have a homogeneous induced colouring.

Conclusion: in this case, an RCAR strategy P with $\mathcal{Y}_P = \mathcal{Y}'$ may exist, but it may be not be easy to find. So in general, for such $\mathcal{Y}'$, we do not know how to efficiently determine if such a P exists.

2. $\mathcal{Y}'$ does have a homogeneous induced colouring.

In this case, we can use (7.4) to compute a candidate q for the RCAR vector. We distinguish three subcases:

2a. If $\mathcal{Y}' \neq \mathcal{Y}$, there may be a message $y \in \mathcal{Y} \setminus \mathcal{Y}'$ for which $\sum_{x \in \mathcal{Y}} q_x > 1$.

This may happen because the described procedure ignores the existence of messages not in $\mathcal{Y}'$. However, the RCAR condition (6.10) puts an inequality constraint on $\sum_{x \in \mathcal{Y}} q_x$ even for messages y with $P(y) = 0$. If the vector q computed by (7.4) does not satisfy this constraint, then q is not an RCAR vector: we chose the wrong $\mathcal{Y}'$.

2b. No solution r of (7.5) achieves $\sum_{y \in \mathcal{Y}} r_y = 1$.

This also means that our choice of $\mathcal{Y}'$ was incorrect.

2c. Otherwise, q is an RCAR vector, and together with r determines an RCAR strategy P .

In this case, we can report success.

In cases 2a and 2b, $\mathcal{Y}'$ has a homogeneous induced colouring but we find that no RCAR strategy P exists with $\mathcal{Y}_P = \mathcal{Y}'$. Then we may face two problems. First, it is not clear how we might choose a different $\mathcal{Y}'$ on which to try the procedure next. For small message structures, it may be feasible to try all candidates. For larger structures, the number of possible choices grows exponentially, and a more efficient way of searching would be needed.

The second problem is that in general, $\mathcal{Y}'$ might not induce a homogeneous colouring even though $\mathcal{Y}$ does. For example, if $\mathcal{Y}$ is the message structure shown in Figure 7.3e, but there is no RCAR strategy P with $\mathcal{Y}_P = \mathcal{Y}$ for our marginal, we have to conclude that the RCAR strategy must have $\mathcal{Y}_P = \{y_1, y_3\}$ (because this is the only other choice of $\mathcal{Y}'$ that satisfies $\bigcup \mathcal{Y}' = \mathcal{X}$). However, this message structure is no longer exchange-connected, and in fact does not have a homogeneous induced colouring, so that we end up in case 1.

In Section 7.5.3, we will see a subclass of matroid games for which the procedure is guaranteed to succeed for the choice $\mathcal{Y}' = \mathcal{Y}$. So for that class of inputs, the procedure discussed here is an efficient algorithm for finding an RCAR strategy (which is worst-case optimal for any loss function by Theorem 7.6).

In Chapter 8, we will see efficient algorithms for graph games and matroid games. The two algorithms in Section 8.3.8 (graphs) and Section 8.5 (matroids) can also be viewed as instances of the computational procedure in this section: both algorithms essentially compute q and r as we did here; then, if $\sum_{y \in \mathcal{Y}} r < 1$, they pick a new set $\mathcal{Y}'$, guided by properties of the linear optimization problem (7.5). The choice of $\mathcal{Y}'$ is such that each new $\mathcal{Y}'$ is a subset of the previous $\mathcal{Y}'$ (i.e. no backtracking is needed), and such that case 2a will never occur.

Case 1 will never occur either for these algorithms: the chosen $\mathcal{Y}'$ will always have a homogeneous induced colouring. This happens for different reasons for the two cases of graph and matroid games. These reasons shed light on what makes graphs and matroids special as message structures of probability updating games, so we conclude this section by giving brief explanations.

For graphs Any connected component of a graph is additionally exchange-connected, and thus induces a homogeneous colouring. While some choices of $\mathcal{Y}'$ may produce a disconnected graph $(\mathcal{X}, \mathcal{Y}')$, each component of this graph will have a homogeneous induced colouring, and the algorithm can be applied to each of these components recursively. (We saw such a decomposition in Example 6.C on page 129, where $\mathcal{Y}$ was exchange-connected, but the strategy that was worst-case optimal for the three standard loss functions used only two disjoint messages.)

For matroids On a matroid game, for any RCAR strategy P , $\mathcal{Y}_P$ determines a homogeneous colouring. (This colouring is not induced in the usual sense, but is uniquely determined by the equalities on $\mathcal{Y}_P$ combined with inequalities for $\mathcal{Y} \setminus \mathcal{Y}_P$; see the proof of Theorem 7.6 for details.) The conditional probabilities $P(x \mid y)$ respect this colouring. This property is stronger than that of graph games, where each component of $\mathcal{Y}_P$ induces a homogeneous colouring, but $\mathcal{Y}_P$ as a whole might not.

7.5.3 Subclasses of matroid games

We now describe a class of games for which a worst-case optimal strategy can be completely computed using the procedure from the previous section, because regardless of the marginal, no messages will need to be discarded.

A message structure $\mathcal{Y}$ is called a *partition matroid* if $\mathcal{X}$ can be partitioned into nonempty sets $S_1, \dots, S_k$ such that $\mathcal{Y}$ contains precisely those subsets of $\mathcal{X}$ that take one element from each of the sets S_i (Oxley, 2011). This class forms a subclass of matroids, so if $\mathcal{Y}$ is a partition matroid, it induces a homogeneous colouring. Using Lemma 7.8, it is easy to see that this colouring is given by the sets S_i . An example of a partition matroid is given in Figure 7.4a; the matroid we saw in Figure 7.3f is not a partition matroid.

	x_1	x_2	x_3	x_4	x_5
y_1	*	—	*	—	—
y_2	*	—	—	*	—
y_3	*	—	—	—	*
y_4	—	*	*	—	—
y_5	—	*	—	*	—
y_6	—	*	—	—	*

(a) Partition matroid but not a sunflower

	x_1	x_2	x_3	x_4	x_5
y_1	*	*	*	—	—
y_2	*	*	—	*	—
y_3	*	*	—	—	*

(b) Sunflower with singleton petals

Figure 7.4: More examples of messages structures and their induced colourings

Because a partition matroid induces a homogeneous colouring, we can perform the procedure described in the previous section to find for each x that $q_x = \sum_{x' \in S_i} p_{x'}$, where S_i is the set containing x . Now a solution for the $P(y)$'s that satisfies $\sum_{y \ni x} P(y) q_x = p_x$ always exists:

$$P(y) = \prod_{x \in y} \frac{p_x}{q_x}.$$

In words, this means that given the true outcome x , it is worst-case optimal for the quizmaster to choose a message by randomly sampling an outcome from each set $S_i \not\ni x$ according to the marginal probabilities conditioned on S_i , and give the message consisting of x and these outcomes. The existence of this strategy shows that, for partition matroid games, the procedure always succeeds in finding a worst-case optimal strategy for the choice $\mathcal{Y}' = \mathcal{Y}$.

What does a message Y generated by this strategy tell the contestant about the true (random) outcome X ? Clearly, it means that if $X \in S_i$ for some i , then X must be the unique outcome in $Y \cap S_i$. Of course, the contestant does not know which of these sets contains X . Write I for the (random) index of the set containing X . Does Y tell the contestant anything about I ? The answer is no: For each index i , regardless of whether $I = i$, the outcome in $Y \cap S_i$ will be randomly distributed according to the marginal p conditioned on S_i , independently of $Y \cap S_j$ for $j \neq i$. This implies that Y is independent of I .

Then for each outcome $x \in Y$, the probability that $X = x$ given message Y equals the probability that $I = i$, where i is the index of the set containing x . These are exactly the probabilities that appear in the RCAR vector q . We know from Theorem 7.6 that the same is true also if the quizmaster is using a worst-case optimal strategy different from the one described above.

For more general message structures, the quizmaster may have to discard a message, so that his worst-case optimal strategy cannot be computed so easily:

Theorem 7.9. *If a game induces a homogeneous colouring but is not a partition matroid, then there exist a marginal and a message $y \in \mathcal{Y}$ such that $P(y) = 0$ for all RCAR strategies P .*

We distinguish one subclass of the class of partition matroid games. A message structure in which the intersection of any two messages is constant is called a *sunflower* (Jukna, 2001). The common intersection is called the *core*, and each set difference between a member and the core is called a *petal*. An example of a *sunflowers with singleton petals* is shown in Figure 7.4b. The Monty Hall game itself (Example 6.B) is another example, and the version of the Monty Hall game with 100 doors from Section 1.2 is a sunflower with 99 petals.

If a message structure is a sunflower with singleton petals, it is a partition matroid: each outcome in the core forms a (singleton) class of the partition, and another class contains all the petals. Among partition matroids, sunflowers can be recognized by the property that all of its colour classes except one are singleton outcomes. For this class of games, the strategy P described above is the *unique* RCAR strategy: a strategy P' with $P'(y) \neq P(y)$ for some $y \in \mathcal{Y}$ would disagree with the unique RCAR vector.

The message structure shown in Figure 7.4a is a partition matroid, but not a sunflower. Because at least two of its colour classes are not singletons, such a message structure contains a cycle of four messages in which neighbouring messages differ by the exchange of one outcome, but the pairs of messages on opposite sides of the cycle differ by two outcomes. (Example 6.G on page 134 is the simplest member of this class of message structures, consisting of just this cycle. In Figure 7.4a, there are three such cycles; one is (y_1, y_2, y_5, y_4) .) For this type of game, the strategy P found above can be modified by increasing $P(y)$ for two messages at opposite sides of the cycle, and decreasing it by the same amount for the other two, leaving the conditionals unchanged. Thus P is not the unique RCAR strategy. In fact, RCAR strategies exist with $P(y) = 0$ for some $y \in \mathcal{Y}$. For such a strategy P , we have $\mathcal{Y}_P \subsetneq \mathcal{Y}$, but we do still have $\sum_{x \in y} q_x = 1$ even for messages y with $P(y) = 0$.

7.6 Discussion and conclusion

7.6.1 Connections to CAR

Recall from Section 6.1 the CAR condition, which characterizes the set of coarsening mechanisms (i.e. quizmaster strategies) for which naive conditioning is

optimal for the contestant. Part of the literature on CAR also addresses the question of whether, for a given message structure, a quizmaster strategy exists that satisfies the CAR condition. A result of Gill et al. (1997) suggests that this is always possible, but Grünwald and Halpern (2003) clarify that this is true only in a very strict sense: for some message structures, a CAR mechanism only exists if some outcomes get probability 0 (so that arguably, we are really dealing with a different message structure). We saw in Example 6.B on page 114 that this is the case in the Monty Hall problem. Further results were found by Jaeger (2005b) and Gill and Grünwald (2008).

Gill and Grünwald introduce the concept of a *uniform multicover* to characterize *rational CAR mechanisms* (and they also show that all CAR mechanisms are finite mixtures of rational CAR mechanisms). This combinatorial structure is closely related to our uniform games, and to games having an induced colouring.

A uniform multicover of $\mathcal{X}$ is a multiset of nonempty subsets $y \subseteq \mathcal{X}$ such that each $x \in \mathcal{X}$ is contained in exactly k such sets (counting multiplicities). Here k is a constant, the same for all outcomes. The translation to our RCAR case involves switching the roles of outcomes and messages. Define the *dual* of a message structure $(\mathcal{X}, \mathcal{Y})$ by $\mathcal{X}' := \mathcal{Y}$ and $\mathcal{Y}' := \{y'_x \mid x \in \mathcal{X}\}$ where $y'_x := \{y \in \mathcal{X}' \mid y \ni x\}$, again allowing multiplicities in $\mathcal{Y}'$. (This operation corresponds to taking the transpose of the incidence matrix of $\mathcal{Y}$.)

It is easy to see that the dual of a uniform game is a uniform multicover. Conversely, if we take the dual of a uniform multicover, respecting its multiplicities (i.e. for each distinct message in the uniform multicover, the dual contains a number of outcomes equal to that message's multiplicity), we obtain a uniform game, though one that may have duplicate messages. If we discard these messages to conform to our usual definition of a game, where multiplicities were not allowed in the set of messages, the result is still a uniform game. But if we then take the dual a second time, we will retrieve a different uniform multicover from the one we started out with. Thus if we want the operation of taking the dual to be its own inverse and want to restrict ourselves to games without duplicate messages, we must also restrict ourselves to uniform multicovers without duplicate outcomes.

Similarly, we may want to forbid duplicate messages on the side of uniform multicovers. If we take the dual of a uniform multicover but ignore its multiplicities, the resulting game may not be uniform, but it will have an induced colouring: if the multiplicity of each message $y \in \mathcal{Y}$ in the uniform multicover is denoted by n_y , then for its dual $(\mathcal{X}', \mathcal{Y}')$, the RCAR vector $q_{x'} = q_y = n_y/k$ is a positive solution to (7.2). For example, in the dual of the uniform multicover shown in Figure 7.5, q would assign probability $2/3$ to the single outcome corresponding to both messages y_1 and y_2 , and $1/3$ to each of the other three outcomes.

A different question is under what conditions on the message structure CAR is *guaranteed* to hold. Grünwald and Halpern (2003, Proposition 4.1) show that this class of message structures is much more limited: this is the case only if $\mathcal{Y}$ is a partition of $\mathcal{X}$. For other message structures, the quizmaster

	x_1	x_2	x_3
y_1	*	*	—
y_2	*	*	—
y_3	*	—	*
y_4	—	*	*
y_5	—	—	*

Figure 7.5: A uniform multicover with a multiple message $y_1 = y_2$. If we take the dual and respect multiplicities, we obtain a uniform game; if we ignore multiplicities, the game obtained is not uniform but does have an induced colouring.

can choose from a nontrivial set of strategies. Among these strategies, the CAR ones and the worst-case optimal ones will in general not coincide, so that naive conditioning may not be worst-case optimal. Whether naive conditioning is a worst-case optimal strategy for given marginal and loss function can easily be checked using our theorems from the previous chapter.

7.6.2 Conclusion

In this chapter, we have seen many classes of message structures, and we found interesting qualitative differences between probability updating games with message structures from different classes. An overview of these results is given in Table 7.1. Possibly the most important of these is the property of loss invariance, shared by graph and matroid games and discussed in Section 7.4.3.

Also, some progress was made in this chapter on the topic of efficiently finding worst-case optimal strategies. First, the results of Section 7.2 can be used to simplify some message structures. If the resulting message structures have homogeneous induced colourings, we may apply the computational procedure of Section 7.5.2; however, this procedure is not guaranteed to produce an answer except in the special case of partition matroids. In Chapter 8, we will see algorithms that efficiently find worst-case optimal strategies for the classes of games for which loss invariance holds: graph and matroid games.

Table 7.1: Results for different classes of message structures

Class	Results	Details
(no induced colouring)	all RCAR strategies must discard a message	page 162
induced colouring	occur as ‘duals’ of uniform multicovers when ignoring multiplicities	page 170
uniform	duals of uniform multicovers without duplicate outcomes	page 170
homogeneous induced colouring	computational procedure to find RCAR strategy is applicable	Section 7.5.2
exchange-connected	induced colouring easy to find	page 163
matroid	RCAR strategies worst-case optimal; loss invariance; efficient algorithm exists; induced colouring very easy to find; $\mathcal{Y}_P$ (with P RCAR) determines a unique homogeneous colouring	Theorem 7.6 Section 7.4.3 Chapter 8 Lemma 7.8 page 167
partition matroid	an RCAR strategy exists that does not discard any messages; computational procedure to find RCAR strategy always succeeds	page 168 page 168
sunflower with singleton petals	there is a unique RCAR strategy (which does not discard messages)	page 169
graph	RCAR strategies worst-case optimal; loss invariance; efficient algorithm exists	Theorem 7.5 Section 7.4.3 Chapter 8

Appendix 7.A Proofs

Proof of Lemma 7.1. For each $y \in \mathcal{Y}$, assume without loss of generality that $y \in \mathcal{Y}_1$. Then observe that the generalized entropies for $\mathcal{G}$ and $\mathcal{G}_1$ are identical on Δ_y ; $P^*(y) > 0$ if and only if $P_1^*(y) > 0$; and $P^*(x | y) = P_1^*(x | y)$ for all $x \in y$. Now the claim follows from Theorem 6.3. $\square$

Proof of Lemma 7.2. We need to show for the P^* and q constructed in the proof that $\sum_{x \in y} q_x \leq 1$ for each $y \in \mathcal{Y}$, with equality if $P^*(y) > 0$, and that P^* satisfies the marginal constraints.

For each $y \in \mathcal{Y}$,

$$\sum_{x \in y} q_x = \sum_{x \in y \setminus \mathcal{X}'} q_x^{\text{out}} + q_{x'}^{\text{out}} \cdot \sum_{x \in y \cap \mathcal{X}'} q_x^{\text{in}} \leq \sum_{x \in y} q_x^{\text{out}} \leq 1. \quad (7.6)$$

If $P^*(y) > 0$, then one of the following holds: if $y \cap \mathcal{X}' = \emptyset$, then $P^{\text{out}}(y) > 0$; if $y \cap \mathcal{X}' \neq \emptyset$, then $P^{\text{in}}(y \cap \mathcal{X}') > 0$ and $P^{\text{out}}(y \setminus \mathcal{X}' \cup \{x'\}) > 0$. In either case, both inequalities in (7.6) are equalities because P^{in} and P^{out} are RCAR strategies.

Now we must show $q_x \cdot \sum_{y \in \mathcal{Y}, y \ni x} P^*(y) = p_x$ for all x . We have for any $x \in \mathcal{X}'$,

$$\begin{aligned} q_x \cdot \sum_{y \in \mathcal{Y}, y \ni x} P^*(y) &= q_{x'}^{\text{out}} \cdot q_x^{\text{in}} \cdot \sum_{y \in \mathcal{Y}, y \ni x} P^{\text{out}}(y \setminus \mathcal{X}' \cup \{x'\}) \cdot P^{\text{in}}(y \cap \mathcal{X}') \\ &= \left(q_{x'}^{\text{out}} \cdot \sum_{\substack{y^{\text{out}} \in \mathcal{Y}^{\text{out}}, \\ y^{\text{out}} \ni x'}} P^{\text{out}}(y^{\text{out}}) \right) \cdot \left(q_x^{\text{in}} \cdot \sum_{\substack{y^{\text{in}} \in \mathcal{Y}^{\text{in}}, \\ y^{\text{in}} \ni x}} P^{\text{in}}(y^{\text{in}}) \right) = p_{x'}^{\text{out}} \cdot p_x^{\text{in}} = p_x. \end{aligned}$$

For $x \notin \mathcal{X}'$ (using that if $P^{\text{out}}(y^{\text{out}}) > 0$ and $x' \in y^{\text{out}}$ for some $y^{\text{out}} \in \mathcal{Y}^{\text{out}}$, then for each $y^{\text{in}} \in \mathcal{Y}^{\text{in}}$ with $P^{\text{in}}(y^{\text{in}}) > 0$, $(y^{\text{out}} \setminus \{x'\}) \cup y^{\text{in}} \in \mathcal{Y}$),

$$\begin{aligned} q_x \cdot \sum_{y \in \mathcal{Y}, y \ni x} P^*(y) &= q_x^{\text{out}} \left(\sum_{\substack{y \in \mathcal{Y}, y \ni x, \\ y \cap \mathcal{X}' = \emptyset}} P^{\text{out}}(y) + \sum_{\substack{y \in \mathcal{Y}, y \ni x, \\ y \cap \mathcal{X}' \neq \emptyset}} P^{\text{out}}(y \setminus \mathcal{X}' \cup \{x'\}) \cdot P^{\text{in}}(y \cap \mathcal{X}') \right) \\ &= q_x^{\text{out}} \cdot \sum_{\substack{y^{\text{out}} \in \mathcal{Y}^{\text{out}}, \\ y^{\text{out}} \ni x}} P^{\text{out}}(y^{\text{out}}) = p_x^{\text{out}} = p_x. \end{aligned}$$

This shows that $P^* \in \mathcal{P}$ and satisfies the RCAR condition. $\square$

Proof of Lemma 7.3. By Theorem 6.3, λ^* is supporting to $H_L \upharpoonright \Delta_{y_1}$ at $P^*(\cdot | y_1)$. Define $\lambda^1 \in \Lambda_{y_1}$ equal to λ^* on y_1 . Then by Lemma 6.4, any $\lambda' \in \Lambda_{y_1}$ with $\lambda' \leq \lambda^1$ obeys $\lambda'_{x_1} = \lambda^1_{x_1}$.

Again by Theorem 6.3, λ^* is dominating to $H_L \upharpoonright \Delta_{y_2}$. Define λ^2 by $\lambda_x^2 = \lambda_x^*$ for $x \in y_1 \cap y_2$, $\lambda_{x_1}^2 = \lambda_{x_2}^*$, and 0 elsewhere. Because L is symmetric between

x_1 and x_2 , $\lambda^2 \in \Lambda_{y_1}$. For $x \in y_1 \cap y_2$, $\lambda_x^1 = \lambda_x^2$. If $\lambda_{x_1}^2 \leq \lambda_{x_1}^1$, then $\lambda^2 \leq \lambda^1$; then we must have $\lambda_{x_1}^2 = \lambda_{x_1}^1$. So $\lambda_{x_1}^2 < \lambda_{x_1}^1$ is impossible, and we find $\lambda_{x_1}^* = \lambda_{x_1}^1 \leq \lambda_{x_1}^2 = \lambda_{x_2}^*$. $\square$

Proof of Lemma 7.4. If P^* is degenerate, then there exist messages y_1, y_2 such that $y_1 \setminus y_2 = \{x_1\}$ and $y_2 \setminus y_1 = \{x_2\}$, with ${}^*P(y_1) > 0$ and $P^*(y_2) > 0$ but $P^*(x_2, y_2) = 0$. Then Lemma 6.2 shows how to construct a worst-case optimal P' with $P'(y_2) = 0$, but otherwise using the same messages that P^* uses. After a finite number of applications of this procedure, we must terminate with a nondegenerate strategy.

Otherwise P^* is nondegenerate. Then for all x_1, x_2 and y_1, y_2 as above with $P^*(y_1) > 0$ and $P^*(y_2) > 0$, we also have $P^*(x_1, y_1) > 0$ and $P^*(x_2, y_2) > 0$. Then we find $\lambda_{x_1}^* = \lambda_{x_2}^*$ by two applications of Lemma 7.3. $\square$

Proof of Theorem 7.5. For graph games, all loss functions are essentially local. We will make this precise by constructing functions f_x , analogous to those in the proof of Theorem 6.10: they have the property that for all $y \in \mathcal{Y}$, a supporting hyperplane to $H_L \upharpoonright \Delta_y$ at $P \in \Delta_y$ is given by λ with $\lambda_x = f_x(P(x))$ for all $x \in y$. (Note that we may not get $f_x(Q(x)) = L(x, Q)$ as in the case of local proper loss functions in the proof of Theorem 6.10, because the hyperplane realized by Q may not be supporting at Q if L is improper.)

For each $x \in \mathcal{X}$, if the only message in which x occurs is $\{x\}$, then $f_x(q)$ is only defined for $q = 1$, where it is $f_x(1) = H_L(e_x)$ (where e_x is the unique element of Δ_y). For other x , pick any message $y \in \mathcal{Y}$ with $y \ni x$ and $|y| = 2$. For all these y , the generalized entropies $H_L \upharpoonright \Delta_y$ are identical copies of the same function, by symmetry of L . For each $q \in [0, 1]$, pick a supporting hyperplane λ to $H_L \upharpoonright \Delta_y$ at the unique $P \in \Delta_y$ with $P(x) = q$, and let $f_x(q) = \lambda_x$. If H_L is not differentiable at P (including when $q \in \{0, 1\}$), we can choose a supporting hyperplane arbitrarily as long as the same one is used to define $f_x(q)$ and $f_{x'}(1 - q)$ wherever $\{x, x'\} \in \mathcal{Y}$. (In particular this means that if a connected component of $\mathcal{Y}$ viewed as a graph contains an odd cycle, $f_x(1/2)$ must take the same value for all x in that component.)

As for local L , each f_x is nonincreasing because H_L is concave, and f_x is strictly decreasing if H_L is strictly concave. The rest of the proof is the same as for Theorem 6.10. $\square$

Proof of Theorem 7.6. We know from Theorem 6.10 that a quizmaster strategy P^* is worst-case optimal for logarithmic loss if and only if it is RCAR, and from Theorem 6.3 that such a P^* exists. Take any such P^* . Let λ be the KT-vector with respect to logarithmic loss, and $\mathcal{Y}_{P^*} = \{y \in \mathcal{Y} \mid P^*(y) > 0\}$. For any pair $y \in \mathcal{Y}_{P^*}, y' \in \mathcal{Y}$, we will show that there exists a bijection π from $y \setminus y'$ to $y' \setminus y$ such that $\lambda_x \leq \lambda_{\pi(x)}$ for all $x \in y \setminus y'$. This follows from Schrijver (2003b, Corollary 39.12a), but here we give a direct proof by induction on $|y' \setminus y|$:

- $|y' \setminus y| = 1$: Apply Lemma 7.3 to $y_1 = y$ and $y_2 = y'$, using that P^* is nondegenerate, to find the required inequality.

- $|y' \setminus y| > 1$: Let $y'_1 = y'$ and pick any $x_1 \in y \setminus y'$. Starting with $i = 1$, apply the basis exchange property on $y \setminus \{x_i\}$ and y'_i to find x'_i (it will be in $y'_i \setminus y \subseteq y'$); then apply it again on $y'_i \setminus \{x'_i\}$ and y to find $x_{i+1} \in y \setminus y'_i$, defining the message $y'_{i+1} = y'_i \setminus \{x'_i\} \cup \{x_{i+1}\}$ (which may not be in $\mathcal{Y}_{P^*}$). Continue until $x_{i+1} = x_1$. Now π defined by $\pi(x_1) = x'_1, \dots, \pi(x_i) = x'_i$ is a bijection from $\{x_1, \dots, x_i\} = (y \cap y'_{i+1}) \setminus y' \subseteq y \setminus y'$ to $\{x'_1, \dots, x'_i\} = y' \setminus y'_{i+1} \subseteq y' \setminus y$ (to see this, note that an element x'_j found in the basis exchange from y is then removed from y'_{j+1} so that it will not be found again; an element x_{j+1} found in the other basis exchange is added to y'_{j+1} with the same result), and for each $1 \leq j \leq i$, applying Lemma 7.3 to y and $y \cup \{x'_j\} \setminus \{x_j\}$ tells us that $\lambda_{x_j} \leq \lambda_{x'_j}$ as required. If $y'_{i+1} = y$, then this is the bijection we are looking for; otherwise, it can be completed by combining it with a bijection from $y \setminus y'_{i+1}$ to $y'_{i+1} \setminus y$, which exists by the induction hypothesis.

If also $y' \in \mathcal{Y}_{P^*}$, a bijection π' from $y' \setminus y$ to $y \setminus y'$ such that $\lambda_{x'} \leq \lambda_{\pi'(x')}$ is found by the same argument. Together, π and π' divide the outcomes in the two sets into disjoint cycles that must all have the same value for λ , defining a colouring of $\mathcal{X}$ that is homogeneous on $\mathcal{Y}_{P^*}$. (Homogeneous colourings are defined in Section 7.5.1.) For logarithmic loss, the RCAR vector q obeys $q_x = e^{-\lambda_x}$, so it must follow the same colouring. Because H_L is strictly concave, the conditionals of P^* must agree with q by Theorem 6.10.

Now take an arbitrary loss function L satisfying the conditions in the theorem, and the same strategy P^* . At an arbitrary message y with $P^*(y) > 0$, choose a supporting hyperplane $\lambda' \in \Delta_y$ to $H_L \upharpoonright \Delta_y$ at $P^*(\cdot | y)$ with $\lambda'_x = \lambda'_{x'}$ wherever x and x' have the same colour: there $P^*(x | y) = P^*(x' | y)$ and L is symmetric between x and x' , so such a supporting hyperplane exists. For all $x, x' \in y$ with $q_x > q_{x'}$ (equivalently, $\lambda_x < \lambda_{x'}$) between which L is symmetric, this λ' satisfies $\lambda'_x \leq \lambda'_{x'}$. (A supporting hyperplane to $H_L \upharpoonright \Delta_y$ at $P^*(\cdot | y)$ with $\lambda'_x > \lambda'_{x'}$ would be lower at $(P^*)^{x_1 \leftrightarrow x_2}(\cdot | y)$ than at $P^*(\cdot | y)$, while by symmetry H_L is the same at those points: a contradiction.)

Because the colouring is homogeneous on $\mathcal{Y}_{P^*}$, the values of λ'_x for $x \in y$ can be copied to all outcomes with the same colour, defining λ' on all of $\mathcal{X}$; for each $y' \in \mathcal{Y}_{P^*}$, λ' defines a supporting hyperplane to $H_L \upharpoonright \Delta_{y'}$ at $P^*(\cdot | y')$.

Also, for each $y' \in \mathcal{Y} \setminus \mathcal{Y}_{P^*}$ and $y \in \mathcal{Y}_{P^*}$, we have that a bijection π exists from $y \setminus y'$ to $y' \setminus y$ such that for all $x \in y \setminus y'$, L is symmetric between x and $\pi(x)$, and $\lambda_x \leq \lambda_{\pi(x)}$; then also $\lambda'_x \leq \lambda'_{\pi(x)}$, so λ' defines a dominating hyperplane to $H_L \upharpoonright \Delta_{y'}$. Thus λ' is a KT-vector certifying that P^* is also worst-case optimal for L .

For the converse: If H_L is strictly concave, the supporting hyperplanes defined by a KT-vector λ' each touch H_L at only one point, so that any worst-case optimal strategy P' for the quizmaster must have $P'(x | y) = q_x$ for all $x \in y$ with $P'(y) > 0$. Therefore any worst-case optimal P' must be RCAR. $\square$

Proof of Theorem 7.7. We will first show how to construct a vector $q \in \mathbf{R}_{>0}^{\mathcal{X}}$

that satisfies $\sum_{x \in y} q_x \leq 1$ for all $y \in \mathcal{Y}$, and for all $x \in \mathcal{X}$, there is a message $y \in \mathcal{Y}$ with $\sum_{x \in y} q_x = 1$. Then we will determine a marginal so that this vector q is the RCAR vector of the game with that marginal. We will additionally find two intersecting messages, both having sum 1, such that q represents the uniform distribution on one, but not on the other.

Two different constructions are given: one for nonuniform and one for uniform games.

If the game is not uniform, let k_2 be the size of the largest message in $\mathcal{Y}$. By connectedness, there exists a message of size less than k_2 that has nonempty intersection with a message of size k_2 . From among such messages, let y_1 be one of maximum size, and let $k_1 < k_2$ be that size. Finally, let y_2 be a message of size k_2 that maximizes $|y_2 \cap y_1|$. Set initial values for q as follows:

$$q_x = \begin{cases} \frac{1}{k_1} & \text{for } x \in y_1; \\ \frac{|y_1 \setminus y_2|}{|y_2 \setminus y_1|} \cdot \frac{1}{k_1} & \text{for } x \in y_2 \setminus y_1; \\ \frac{1}{|y_2 \setminus y_1|} \cdot \frac{1}{k_1} & \text{otherwise.} \end{cases}$$

Note that the three cases of q_x are listed in nonincreasing order. Now $\sum_{x \in y_1} q_x = \sum_{x \in y_2} q_x = 1$, while $\sum_{x \in y} q_x \leq 1$ for general $y \in \mathcal{Y}$: $\max_x q_x = 1/k_1$, so a message $y \in \mathcal{Y}$ with $|y| \leq k_1$ will have sum at most 1; a message with $|y| = k_2$ will share no more outcomes with y_1 than y_2 does and thus cannot have a larger sum; and because a message with $k_1 < |y| < k_2$ has empty intersection with y_2 , the $k_1 - 1$ largest elements of $(q_x)_{x \in y}$ sum to at most $(k_1 - 1)/k_1$, while the fewer than $|y_2 \setminus y_1|$ remaining elements all equal $1/(|y_2 \setminus y_1| \cdot k_1)$ and hence sum to less than $1/k_1$.

A greedy algorithm that repeatedly increments some q_x until none can be increased further, while maintaining the inequality $\sum_{x \in y} q_x \leq 1$ on each y , will terminate with a q satisfying the conditions stated at the beginning of the proof. This q will be unchanged and thus still be uniform on y_1 , while on the intersecting message y_2 , q also still sums to 1 but is not uniform.

For the case of uniform games, the construction is similar. Let k be the size of the game's messages. By Oxley (2011, Corollary 2.1.5), a nonempty family of sets $\mathcal{Y}$ is the collection of bases of a matroid if and only if for all $y_1, y_2 \in \mathcal{Y}$ and $x_2 \in y_2 \setminus y_1$,

$$y_1 \cup \{x_2\} \setminus \{x_1\} \in \mathcal{Y} \text{ for some } x_1 \in y_1 \setminus y_2. \quad (7.7)$$

Because our $\mathcal{Y}$ is not a matroid, it follows that there exist $y_1, y_2 \in \mathcal{Y}$ and $x_2 \in y_2 \setminus y_1$ for which no corresponding x_1 exists. For $k \geq 3$ (which holds because the game we consider is not a graph game), we claim something stronger: that there exist y_1, y_2, x_2 as above with the additional property that y_1 and y_2 intersect. The proof of this claim is below.

Using such y_1 and x_2 and some $0 < \epsilon < 1/k$, initialize q as follows:

$$q_x = \begin{cases} \frac{1}{k} & \text{for } x \in y_1; \\ \frac{1}{k} + \epsilon & \text{for } x = x_2; \\ \frac{1}{k} - \epsilon & \text{otherwise.} \end{cases}$$

Because any message containing x_2 also contains at least one other outcome not in y_1 , we again have $\sum_{x \in y} q_x \leq 1$ for all $y \in \mathcal{Y}$.

For $k \geq 3$, the initial q has the property that the set of outcomes x for which q_x cannot be increased further (we call these outcomes *maximized*) is connected by messages y with $\sum_{x \in y} q_x = 1$ (that is, the maximized outcomes cannot be partitioned into two nonempty sets such that each sum-1 message is contained in one of these sets); this is because y_1 has sum 1, and any other message with sum 1 must intersect y_1 . (For $k = 2$, this would not be the case: the only messages having sum 1 would be y_1 and all messages that contain x_2 , but y_1 would not intersect any of these.) We can have the greedy algorithm maintain this as an invariant: Because the game is connected, there is always a message partially in the set of maximized outcomes and partially outside. We call such a message a *crossing* message. Each round, we pick an outcome x that is not maximized yet and is contained in a crossing message; if x_2 is not maximized, we always pick $x = x_2$ (using that it is contained in the crossing message y_2). The tightest constraint on increasing q_x will come from a crossing message, because for any non-crossing message $y \ni x$, we have $\sum_{x' \in y \setminus \{x\}} q_{x'} = (k - 1)(1/k - \epsilon)$, which is the smallest possible value of this sum. So increasing q_x as much as possible will cause a crossing message to get sum equal to 1. This message connects x to the set of previously maximized outcomes, and any other outcomes that were maximized by this increment must be contained in some message that also contains x .

When the greedy algorithm terminates, q will still be uniform on y_1 , while there will be another message on which q sums to one but is not uniform. (This may not be y_2 , which may not have sum 1.) Because all outcomes are connected by sum-1 messages, we can also find a pair of intersecting messages, one of which is uniform and one of which is not. Use these two messages as y_1 and y_2 in the sequel.

Having found, for both nonuniform and uniform games, a vector q and messages y_1 and y_2 as described above, we let strategy P be RCAR with vector q and $P(y)$ uniform on $\{y \mid \sum_{x \in y} q_x = 1\}$. This P is a worst-case optimal strategy for the game with logarithmic loss and marginal $p_x = \sum_{y \ni x} P(x, y)$, and q is its unique RCAR vector.

We will show that P is not worst-case optimal for the game with the same marginal and Brier loss. Brier loss is proper and continuous, so by Theorem 6.9, $L(x, P(\cdot \mid y_1)) = L(x, P(\cdot \mid y_2))$ for worst-case optimal P . These are squared Euclidean distances from a vertex of the simplex to the predicted distribution. However, the equality will not hold for P :

Among all predictions in Δ_{y_i} with $Q(x) = q_x$ for each $x \in y_1 \cap y_2$ (this set of predictions is the intersection of Δ_{y_i} with an affine subspace), the squared Euclidean distance $L(x, Q)$ between such Q and given vertex $x \in y_1 \cap y_2$ is uniquely minimized by Q uniform on the outcomes not in $y_1 \cap y_2$ (this is the orthogonal projection of the vertex onto that subspace). For a uniform game, $P(\cdot \mid y_1)$ is uniform and thus $L(x, P(\cdot \mid y_1))$ equals this minimum; $P(\cdot \mid y_1)$ differs from the uniform distribution at some outcomes not in $y_1 \cap y_2$ and thus $L(x, P(\cdot \mid y_2))$ is larger than the minimum.

For a nonuniform game, $P(\cdot \mid y_1)$ is uniform on $y_1 \setminus y_2$ and $P(\cdot \mid y_2)$ is uniform on $y_2 \setminus y_1$, so both minimize the distance to the vertex in their respective subspaces. However, the subspace for y_2 is isomorphic to a subspace contained in the subspace for y_1 and not containing $P(\cdot \mid y_1)$. Therefore $L(x, P(\cdot \mid y_1)) < L(x, P(\cdot \mid y_2))$. $\square$

Proof of claim. Suppose for a contradiction that any pair of intersecting messages y, y' obeys the above exchange property (7.7) for all $x' \in y' \setminus y$. Let y_1, y_2 be two messages that fail (7.7) for some outcome $x_2 \in y_2 \setminus y_1$; it follows from our assumption that they are disjoint. Because $\mathcal{Y}$ is connected, there exists a sequence of messages starting with y_1 and ending with y_2 in which adjacent messages intersect. Using (7.7), we can extend this sequence to one where adjacent messages differ by the exchange of one outcome: given intersecting $y, y'' \in \mathcal{Y}$ with $d := |y'' \setminus y| > 1$, we find $y' \in \mathcal{Y}$ with $|y' \setminus y| = 1$ and $|y'' \setminus y'| = d - 1$. Write the entire sequence as $y^0 = y_1, y^1, \dots, y^n = y_2$.

We have $n \geq k$, because $n < k$ would imply that $y_1 \cap y_2 \neq \emptyset$. If $n > k$, we can find a shorter sequence as follows: pick $0 \leq i < j \leq n - k$ for which $y^i \cap y^{j+1} \neq \emptyset$; this holds if $j + 1 - i < k$, so such i, j can always be found if $k \geq 3$. Let x' be the unique outcome in $y^{j+1} \setminus y^j$.

- If $x' \notin y^{j+k}$ (intuitively, adding x' leads us on a detour that can be avoided when going to y^{j+k}): In each of the k exchange steps from y^j to y^{j+k} , one outcome was removed. One of those outcomes was x' , which is not in y^j , so at most $k - 1$ outcomes from y^j were removed. Thus y^j and y^{j+k} intersect, and a shorter path between them can be found using (7.7).
- If $x' \in y^{j+k}$ and $x' \in y^i$ (removing x' is the start of a detour): We can use (7.7) to find a shorter path between y^i and y^{j+k} .
- If $x' \in y^{j+k}$ but $x' \notin y^i$ (adding x' is apparently useful, but can be done sooner): Apply (7.7) to messages y^i and y^{j+1} (which intersect) and outcome x' (which is in y^{j+1} but not in y^i) to find a message y' that is one step away from y^i and contains x' . From y' , we can find a path to y^{j+k} by (7.7) taking fewer than k steps. Thus we can get from y^i to y^{j+k} in at most k steps.

Thus we can always find a sequence with $n = k$.

Given such a sequence $y^0, y^1, \dots, y^n$, we will now show a contradiction with the assumption that $y_1 = y^0$ and $y_2 = y^n$ fail (7.7) by showing that for any $x_2 \in y_2$, a message exists that differs from y_1 by adding x_2 and removing one other outcome. If $x_2 \in y^1$, then y^1 is such a message and we are done. Otherwise, we can apply (7.7) to y^1 and $x_2 \in y_2$ to find a message y' containing x_2 ; because $k \geq 3$, this message still intersects y_1 , so applying (7.7) to y_1 and $x_2 \in y'$ gives the message we are looking for. This shows by contradiction that if a connected uniform game with $k \geq 3$ is not a matroid game, there exists a pair of intersecting messages y_1, y_2 and an outcome $x_2 \in y_2 \setminus y_1$ that do not satisfy (7.7). $\square$

Proof of Lemma 7.8. ($2 \Leftarrow 3$) Two elements $x_1 \neq x_2$ of $\mathcal{X}$ are in the same 2-connected component if and only if there is a *circuit* (minimal dependent set) containing both. Since a basis is a maximal dependent set, $y_1 \cup y_2$ is independent. Find a circuit $C \subseteq y_1 \cup y_2$; this circuit contains both x_1 and x_2 , as otherwise it would be contained in a basis and thus independent.

($2 \Rightarrow 3$) Let C be a circuit with $\{x_1, x_2\} \subseteq C$; our goal is to find the bases y_1, y_2 , which we will do iteratively. Let y_1 be a basis containing the independent set $C \setminus \{x_2\}$, and y_2 a basis containing $C \setminus \{x_1\}$. While $y_1 \setminus \{x_1\} \neq y_2 \setminus \{x_2\}$, pick any $x'_1 \in y_1 \setminus (y_2 \cup \{x_1\})$ and use basis exchange to find a basis $y' = (y_1 \setminus \{x'_1\}) \cup \{x'_2\}$ for some $x'_2 \in y_2 \setminus y_1$. Note that $x'_2 \neq x_2$, as that would result in $C \subseteq y'$. Replace y_1 by y' and repeat until $y_1 \setminus \{x_1\} = y_2 \setminus \{x_2\}$. This process terminates, as the set difference becomes smaller with each step.

($1 \Leftrightarrow 3$) For exchange-connected message structures, the colour classes are the equivalence classes of the transitive reflexive closure of the relation on $\mathcal{X}$ stated in point 3. For matroids, the equivalence of points 2 and 3 shows that this relation is already transitive. Thus for all $x_1 \neq x_2$, points 1 and 3 are equivalent.

Finally, a matroid is equal to the direct sum of its 2-connected components (Oxley, 2011, Corollary 4.2.9), which shows that the 2-connected components (or equivalently, the classes of the induced colouring) are modules. $\square$

Proof of Theorem 7.9. The proof technique is similar to the one used to prove Theorem 7.7: we construct a marginal with the required property by first finding a vector q that is the RCAR vector for some game with the given message structure.

We distinguish two cases. If there exists $y' \subseteq \mathcal{X}$ that is consistent with the homogeneous induced colouring but $y' \notin \mathcal{Y}$, then pick $0 < \epsilon < 1/(k(k-1))$ and set initial values for q as follows:

$$q_x = \begin{cases} \frac{1}{k} + \epsilon & \text{for } x \in y'; \\ \frac{1}{k} - (k-1)\epsilon & \text{otherwise.} \end{cases}$$

Because each message contains at least one outcome with $q_x = 1/k - (k-1)\epsilon$, we have $\sum_{x \in y} q_x \leq 1$ for all $y \in \mathcal{Y}$.

Otherwise there must exist a colour class $C \subseteq \mathcal{X}$ for which the number of outcomes of this colour occurring in a message is at least two. (If all colours occur exactly once in each message and all $y' \subseteq \mathcal{X}$ consistent with this colouring are $y' \in \mathcal{Y}$, then $\mathcal{Y}$ is a partition matroid.) Then pick any $x^+ \in C$ and $0 < \epsilon < 1/k$, and initialize q according to

$$q_x = \begin{cases} \frac{1}{k} + \epsilon & \text{for } x = x^+; \\ \frac{1}{k} - \epsilon & \text{for } x \in C \text{ but } x \neq x^+; \\ \frac{1}{k} & \text{otherwise.} \end{cases}$$

Again we see $\sum_{x \in y} q_x \leq 1$ for all $y \in \mathcal{Y}$.

We apply the same greedy algorithm we used in the proof of Theorem 7.7: repeatedly increase q_x for some x until none can be increased further, maintaining $\sum_{x \in y} q_x \leq 1$ for all $y \in \mathcal{Y}$. For a vector q obtained by this algorithm, let P be the joint distribution on x, y with $x \in y$ for which the marginal $P(y)$ is uniform on $\{y \in \mathcal{Y} \mid \sum_{x \in y} q_x = 1\}$ and for which $P(x \mid y) = q_x$ for all $x \in y$. This P is an RCAR strategy for the game with marginal $p_x = \sum_{y \ni x} P(x, y)$, and q is the RCAR vector.

In the first case, there must exist some $x^- \in \mathcal{X}$ with $q_{x^-} \leq 1/k$. Let C be the colour class containing x^- , and let x^+ be the unique outcome in $C \cap y'$. In the second case, there must exist some $x^- \in C$ with $q_{x^-} \leq 1/k$. Thus in either case, we have two outcomes x^- and x^+ of the same colour C but with $q_{x^-} \leq 1/k < 1/k + \epsilon \leq q_{x^+}$. Because this contradicts the definition of an induced colouring, there must be a message for which q violates the equality (7.2). This message must be discarded by any RCAR strategy for this game. $\square$