Blackwell's Informativeness Theorem Using Diagrams

Henrique de Oliveira[1,*]

[a]*Department of Economics, Pennsylvania State University, University Park, PA 16802*

Abstract

This paper gives a simple proof of Blackwell's theorem on the ranking of information structures. The proof extends naturally to environments where information arrives over time (leading to the notion of adapted garbling) and environments where information is diffused among multiple players (leading to the notion of independent garbling).

Keywords: Blackwell's Theorem, Information, Garbling, Category Theory

1. Introduction

An agent faces a *decision problem under uncertainty*, whose payoff $u(a, \omega)$ depends on her action $a \in A$ and the state of the world $\omega \in \Omega$.[2] The agent does not know ω but observes an informative random signal $s \in S$, drawn according to the *information structure* $\sigma : \Omega \to \Delta(S)$, which specifies the conditional probability $\sigma(s|\omega)$ of observing signal s when the state is ω.

Blackwell (1951, 1953) proved that three different methods to rank information structures generate the same order. The first ranking comes from a notion of "adding noise": Say that σ' *is a garbling of* σ if an agent who knows σ could replicate σ' by randomly drawing a signal $s' \in S'$ after each observation of $s \in S$, i.e. there exists a function $\gamma : S \to \Delta(S')$ (the garbling)

*

Email address: `henrique@psu.edu` (Henrique de Oliveira)

[1]Part of the research that led to this paper was conducted while visiting Princeton University. I thank Tomasz Strzalecki for a stimulating conversation that originated this project. I also thank Nageeb Ali, Stephen Morris, Wolfgang Pesendorfer, Philipp Sadowski, Todd Sarver and Asher Wolinsky for their helpful comments and discussions.

[2]For simplicity, all sets will be assumed to be finite.

such that

$$\sigma'(s'|\omega) = \sum_{s\in S} \gamma(s'|s)\,\sigma(s|\omega).$$

The second ranking comes from a notion of feasibility: Given a set of actions A and an information structure σ, a mixed strategy $\alpha : S \to \Delta(A)$ induces a distribution over actions conditional on ω. We can then rank σ and σ' according to which yields the larger set of feasible conditional distributions of actions. The third ranking comes from thinking in terms of expected utility: Say that σ is more informative than σ' if every Bayesian agent, facing any decision problem, can obtain a higher expected utility using σ than by using σ'. Blackwell's theorem can then be stated thus:

Theorem 1. *The following statements are equivalent:*

1. *σ' is a garbling of σ;*
2. *For any set of actions A, the set of conditional distributions over actions that are feasible under σ contains those that are feasible under σ';*
3. *Every Bayesian agent prefers σ to σ', for any possible decision problem.*

In this paper, I provide a proof of Blackwell's Theorem which uses simple categorical properties of stochastic maps to show the equivalence of parts 1 and 2. Blackwell's original proof was rather involved and there have been attempts to simplify it (Crémer (1982), Bielinska-Kwapisz (2003), Leshno and Spector (1992)). These papers represent information structures and garblings as matrices and use matrix algebra to derive the result. Focusing on the categorical properties of stochastic maps avoids cluttering notation that comes from working with matrices. These categorical properties underlie some of the arguments in Lehrer et al. (2013), but I believe that this is the first paper to explicitly state these properties in connection to Blackwell's Theorem.

The ideas present in the proof can be useful to think about information in other contexts as well. To illustrate this, I show two other applications: to information that arrives over time, reformulating a result of Greenshtein (1996), and to information in strategic environments, deriving the concept of Independent Garbling (Lehrer et al., 2010).

2. The category of stochastic maps

Given a finite set X, let $\Delta(X)$ denote the set of probability distributions over X. A *stochastic map* between two finite sets X and Y is a function $\alpha : X \to \Delta(Y)$. Given two stochastic maps $\alpha : X \to \Delta(Y)$ and $\beta : Y \to \Delta(Z)$, their *composition* is the stochastic map $\beta \circ \alpha : X \to \Delta(Z)$ defined by

$$\beta \circ \alpha (z|x) = \sum_{y \in Y} \beta(z|y)\, \alpha(y|x)\,.$$

This composition operation is associative: given another stochastic map $\gamma : Z \to \Delta(W)$, we can write

$$(\gamma \circ \beta) \circ \alpha = \gamma \circ (\beta \circ \alpha) = \sum_{z \in Z} \sum_{y \in Y} \gamma(w|z)\, \beta(z|y)\, \alpha(y|x)$$

and the order of summation does not matter.

A special stochastic map is the *identity map* $id_Y : Y \to \Delta(Y)$, which takes a point $y \in Y$ to the measure $\delta_y \in \Delta(Y)$ that puts probability 1 on y. The identity map has the property that given any maps $\alpha : X \to \Delta(Y)$ and $\beta : Y \to \Delta(Z)$, we have $id_Y \circ \alpha = \alpha$ and $\beta \circ id_Y = \beta$. The existence of an identity, together with the associativity property of composition, formally make stochastic maps a category[3] As is typical in category theory, we will often represent stochastic maps as diagrams of arrows between finite sets, and say that such *a diagram commutes* if any two paths with the same beginning and end define the same stochastic map through composition. For example, saying that the diagram

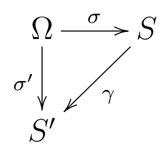
$$\begin{array}{ccc} A & \xrightarrow{\alpha} & B \\ {\scriptstyle \alpha'}\downarrow & & \downarrow{\scriptstyle \beta} \\ B' & \xrightarrow{\beta'} & C \end{array}$$

commutes is to say that $\beta \circ \alpha = \beta' \circ \alpha'$.

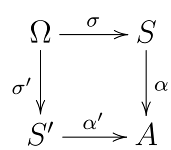
[3]A *category* is a collection of objects (X, Y, Z etc), arrows between objects (α,β,γ etc) and a binary operation $\circ$, which takes two arrows $\alpha : X \to Y$ and $\beta : Y \to Z$ into a new arrow $\beta \circ \alpha : X \to Z$, satisfying two properties: (1) Every object Y has an *identity arrow* $id_Y : Y \to Y$ such that for any $\alpha : X \to Y$ we have $id_Y \circ \alpha = \alpha$ and for any $\beta : Y \to Z$ we have $\beta \circ id_Y = \beta$ and (2) $\circ$ is associative (Mac Lane, 1978, page 7).

3. Garblings and strategies

Information structures, garblings and strategies are all stochastic maps. Thus, Condition (1) in Theorem 1 can be succinctly stated as the existence of a stochastic map γ such that $\sigma' = \gamma \circ \sigma$. In a diagram,



Likewise, the conditional distribution over actions induced by a strategy α is simply $\alpha \circ \sigma$. Condition (2) in the statement of Blackwell's theorem can be stated in a diagram as "for every α' there exists an α such that the following diagram commutes:"



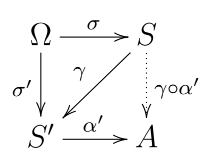
The expected utility of a strategy only depends on the conditional probability over actions, for it can be written as

$$\sum_{\omega} \left(\sum_{a \in A} u(a, \omega)\, \alpha \circ \sigma(a|\omega) \right) p(\omega),$$

where $u : A \times \Omega \to \mathbb{R}$ is the utility function and p is the prior.

4. Proof of Blackwell's Theorem

(1)⇒(2). We can see this by simply completing the diagram that defines a garbling by defining $\alpha = \gamma \circ \alpha'$:



This diagram commutes, so we have (2).

(2)⇒(1). Let α' be the identity map for the signals of σ', $Id_{S'}$. Condition (2) tells us that there exists an α that makes the following diagram commutes

$$\begin{array}{ccc} \Omega & \xrightarrow{\sigma} & S \\ {\scriptstyle \sigma'}\downarrow & & \downarrow{\scriptstyle \alpha} \\ S' & \xrightarrow{Id_{S'}} & S' \end{array}$$

But note that in that case we have $\sigma' = Id_{S'} \circ \sigma' = \alpha \circ \sigma$, so that α is a garbling from σ to σ'.

(2)⇒(3). Fix a set of actions A and a utility function $u : A \times \Omega \to \mathbb{R}$. Let Λ_σ be the set of all conditional probability over actions that are feasible under σ

$$\Lambda_\sigma = \{\alpha \circ \sigma | \alpha : S \to \Delta(A)\}.$$

We can write the expected utility of an agent as

$$\max_{\lambda \in \Lambda_\sigma} \sum_\omega \left(\sum_{a \in A} u(a, \omega) \lambda(a|\omega) \right) p(\omega).$$

Clearly, if $\Lambda_\sigma \supseteq \Lambda_{\sigma'}$ the agent can obtain higher expected utility using σ than using σ'.

(3)⇒(2). We can prove the contrapositive: If (2) is not true, then there exists a $\lambda' \in \Lambda_{\sigma'}$ such that $\lambda' \notin \Lambda_\sigma$. Seeing Λ_σ as a subset of $\mathbb{R}^{\Omega \times A}$, it is easy to show that it is compact and convex. If A has at least two actions, this means that there must exist a separating hyperplane orthogonal to a vector $v \in \mathbb{R}^{\Omega \times A}$, such that

$$\sum_{\omega, a} v(a, \omega) \lambda(a|\omega) < \sum_{\omega, a} v(a, \omega) \lambda'(a|\omega) \quad \forall \lambda \in \Lambda_\sigma$$

Thus, a Bayesian agent with utility v and uniform prior can obtain strictly higher utility using σ' than using any strategy under σ.

5. Dynamic Informativeness

Now consider an agent whose information arrives over time. Let $t = 1, \ldots, T$ denote the time periods. At each period, the agent observes a new signal $s_t \in S_t$ and then takes an action $a_t \in A_t$. A *dynamic information structure* $\sigma : \Omega \to \Delta(S_1 \times \cdots \times S_T)$ specifies the probability of all sequences of signals, given a state of the world. A *dynamic decision problem* is given by the sets $A_1, \dots, A_T$ and a utility function $u : A_1 \times \cdots \times A_T \times \Omega \to \mathbb{R}$.

The agent can condition the choice of $a_t \in A_t$ on the realization of all past signals $s_1, \dots, s_t$ but not on the signals $s_{t+1}, \dots, s_T$, which have not been revealed yet. Hence a *strategy* of the agent specifies a (possibly random) action to be taken after each history of past signals and past actions, that is, it is a collection of mappings $(\alpha_t)_{t=1}^{T}$ where $\alpha_1 : S_1 \to \Delta(A_1)$, $\alpha_2 : S_1 \times S_2 \times A_1 \to \Delta(A_2)$ etc. For our purposes, it will be more convenient to represent strategies in an alternative way, which we explain below. To simplify notation, let $S^t = S_1 \times \cdots \times S_t$, $A^t = A_1 \times \cdots \times A_t$ etc for all $t = 1, \dots, T$.

Definition 2. A mapping $\alpha : S^T \to \Delta\left(A^T\right)$ is *adapted* if its marginal distribution on A^t depends only on S^t, that is, if there exist maps $\alpha^t : S^t \to \Delta(A^t)$ such that

$$\alpha^t (a_1, \dots, a_t | s_1, \dots, s_t) = \sum_{a_{t+1}, \dots, a_T} \alpha (a_1, \dots, a_t, a_{t+1}, \dots, a_T | s_1, \dots, s_T)$$

for all $t = 1, \dots, T$, $(s_1, \dots s_T) \in S^T$, and $(a_1, \dots, a_T) \in A^t$.

Lemma 3. *Every strategy induces a unique adapted mapping. Every adapted mapping is induced by some strategy.*

Proof. Given a strategy $(\alpha_t)_{t=1}^{T}$, let $\alpha^1 = \alpha_1$ and define inductively

$$\alpha^t (a_1, \dots, a_t | s_1, \dots, s_t) = \alpha^{t-1} (a_1, \dots, a_{t-1} | s_1, \dots, s_{t-1}) \, \alpha_t (a_t | s_1, \dots, s_t, a_1, \dots, a_{t-1}) .$$

We can prove by induction that α^t is adapted for every t. Indeed, suppose that α^{t-1} is adapted. The marginal of α^t on A^{t-1} is simply α^{t-1}, which depends only on $s_1, \dots, s_{t-1}$. The marginal of α^t on A^r for $r < t-1$ is the same as the marginal of α^{t-1} on A^r and therefore depends only on S^r, since α^{t-1} is adapted. Therefore α^t is adapted.

Now suppose that we have an adapted mapping $\alpha : S^T \to \Delta\left(A^T\right)$. Given the mappings corresponding to the marginals $\alpha^t : S^t \to \Delta(A^t)$, let $\alpha_t : S^t \times A^{t-1} \to \Delta(A_t)$ give a conditional probability of a_t given $a_1, \dots, a_{t-1}$ and $s_1, \dots, s_t$. [4] Then we can show that $\alpha^t = \alpha^{t-1} \alpha_t$ as before, so that the strategy $(\alpha_t)_{t=1}^{T}$ induces the adapted mapping α^T. □

[4]Bayes rule must be respected when the probability of $a_1, \dots, a_{t-1}$ given $s_1, \dots, s_t$ is greater than zero. When it is zero, α_t may be defined arbitrarily.

Adapted mappings can be conveniently represented using a commutative diagram. Let $\pi_t : S^{t+1} \to S^t$ and $\rho_t : A^{t+1} \to A^t$ be the projection mappings (these can be regarded as stochastic mappings that put probability one on a single value). Given a stochastic mapping $\alpha^{t+1} : S^{t+1} \to \Delta\left(A^{t+1}\right)$, the composition $\rho_t \circ \alpha^{t+1} : S^{t+1} \to \Delta\left(A^t\right)$ gives precisely the marginal distribution of α^{t+1} on A^t. If α^{t+1} is adapted, then that marginal depends only on S^t: there exists a mapping $\alpha^t : S^t \to \Delta\left(A^t\right)$ such that $\alpha^t \circ \pi_t = \rho_t \circ \alpha^{t+1}$. Thus a mapping $\alpha : S^T \to \Delta\left(A^T\right)$ is adapted if there exist mappings α^t that make the following diagram commute:

$$\begin{array}{ccccccc} S^1 & \xleftarrow{\pi_1} & S^2 & \xleftarrow{\pi_2} & \cdots & \xleftarrow{\pi_T} & S^T \\ {\scriptstyle \alpha^1}\downarrow & & {\scriptstyle \alpha^2}\downarrow & & & & {\scriptstyle \alpha}\downarrow \\ A^1 & \xleftarrow{\rho_1} & A^2 & \xleftarrow{\rho_2} & \cdots & \xleftarrow{\rho_T} & A^T \end{array}$$

Lemma 4. *The adapted mappings form a category under the operation of composition of stochastic mappings.*

Proof. Let $\alpha : S^T \to \Delta\left(A^T\right)$ and $\beta : A^T \to \Delta\left(B^T\right)$ be adapted mappings. The only nontrivial step here is to show that their composition $\beta \circ \alpha : S^T \to \Delta\left(B^T\right)$ is also an adapted mapping. To see this, we can write the diagram that represents their composition:

$$\begin{array}{ccccccc} S^1 & \xleftarrow{\pi_1} & S^2 & \xleftarrow{\pi_2} & \cdots & \xleftarrow{\pi_T} & S^T \\ {\scriptstyle \alpha^1}\downarrow & & {\scriptstyle \alpha^2}\downarrow & & & & {\scriptstyle \alpha}\downarrow \\ A^1 & \xleftarrow{\rho_1} & A^2 & \xleftarrow{\rho_2} & \cdots & \xleftarrow{\rho_T} & A^T \\ {\scriptstyle \beta^1}\downarrow & & {\scriptstyle \beta^2}\downarrow & & & & {\scriptstyle \beta}\downarrow \\ B^1 & \xleftarrow{\lambda_1} & B^2 & \xleftarrow{\lambda_2} & \cdots & \xleftarrow{\lambda_T} & B^T \end{array}$$

It is easy to see that this diagram commutes. For example,

$$\lambda_1 \circ \beta_2 \circ \alpha_2 = \beta_1 \circ \rho_1 \circ \alpha_2 = \beta_1 \circ \alpha_1 \circ \pi_1.$$

This means that $\beta \circ \alpha$ is an adapted mapping, with the mappings $\beta_t \circ \alpha_t$ corresponding to its marginals. □

In the proof of Theorem 1, a key property used was that garblings and strategies were the same kind of mathematical objects. There, they were stochastic mappings; here, strategies must be adapted mappings. This suggests that the right notion of garbling for the present dynamic setting is one where the garbling is required to be an adapted mapping. The following theorem, due to Greenshtein, confirms this suggestion.

Theorem 5. *The following statements are equivalent:*

1. *σ' is an adapted garbling of σ;*

2. *For any sets of actions $A_1, \ldots, A_T$, the set of conditional distributions over $A_1 \times \cdots \times A_T$ that are feasible under σ contains those that are feasible under σ';*

3. *Every Bayesian agent prefers σ to σ', for any possible dynamic decision problem.*

The proof follows exactly the same lines as the proof of Theorem 1 so it is omitted.

The condition that σ' be an adapted garbling of σ is not always obvious. For example it may be that the marginal distribution σ^t is more informative in the sense of Blackwell than σ'^t for every t, yet σ' is not an adapted garbling of σ. This can be the case even if the information structures represent the arrival of independent signals (see Greenshtein (1996) for an example).

6. Strategic Informativeness

Now consider two agents, each receiving a signal about a common state. We can represent such a *joint information structure* as a stochastic mapping $\sigma : \Omega \to \Delta(S_1 \times S_2)$, where $s_i \in S_i$ is the signal observed by agent i. After seeing their signal, each agent takes an action from a set A_i. Thus, a strategy for agent i is a stochastic mapping $\alpha_i : S_i \to \Delta(A_i)$; the *strategy profile* is a stochastic mapping $\alpha \equiv \alpha_1 \otimes \alpha_2 : S_1 \times S_2 \to \Delta(A_1 \times A_2)$, obtained by taking the independent product of each player's strategy. The independence can be understood as an assumption that all the information the players have available to them is expressed in the joint information structure σ.

Together, a joint information structure and a strategy profile induce a conditional distribution over actions $\alpha \circ \sigma : \Omega \to \Delta(A_1 \times A_2)$. When is it that σ permits more conditional distributions over actions than an alternative joint information structure σ'? Again, the key property here is that garblings and strategies be the same kind of mathematical objects. In this case, we will say that $\gamma : S_1 \times S_2 \to \Delta(S_1' \times S_2')$ is an *independent garbling* if there exist $\gamma_1 : S_1 \to \Delta(S_1')$ and $\gamma_2 : S_2 \to \Delta(S_2')$ such that γ is the independent product of γ_1 and γ_2, i.e.

$$\gamma(s_1', s_2' | s_1, s_2) = \gamma_1(s_1' | s_1)\, \gamma_2(s_2' | s_2).$$

That this is the right notion of garbling is confirmed by the following result.

Theorem 6. *The following statements are equivalent*

1. *σ' is an independent garbling of σ;*
2. *For any sets of actions A_1 and A_2, the set of conditional distributions over $A_1 \times A_2$ that are feasible under σ contains those that are feasible under σ'.*

The proof again follows easily the first half of the proof of Theorem 1 once we note that independent garblings form a category.

Notice that this theorem, unlike Theorems 1 and 5, makes no mention of preferences. In order to do that, we would have to specify a game-theoretic solution concept under which players would prefer σ to σ'. This comes with some difficulties, e.g. multiplicity of equilibria, but most importantly it may be that having more feasible actions actually hurt players (they may value the commitment power that comes from having less information). Some proposed solutions involve restricting the class of games under consideration (so that there's no value for commitment) (Lehrer et al., 2010, Pęski, 2008) or ranking information structures according to the set of equilibria they generate (Bergemann and Morris, 2015, Cherry and Smith, 2012, Gossner, 2000).

Dirk Bergemann and Stephen Morris. Bayes correlated equilibrium and the comparison of information structures. *Theoretical Economics*, 2015.

Agnieszka Bielinska-Kwapisz. Sufficiency in Blackwell's theorem. *Mathematical Social Sciences*, 46(1):21–25, 2003.

D. Blackwell. Comparison of experiments. In *Second Berkeley Symposium on Mathematical Statistics and Probability*, volume 1, pages 93–102, 1951.

D. Blackwell. Equivalent comparisons of experiments. *The Annals of Mathematical Statistics*, pages 265–272, 1953.

Josh Cherry and Lones Smith. Strategically valuable information. 2012.

Jacques Crémer. A simple proof of Blackwell's "comparison of experiments" theorem. *Journal of Economic Theory*, 27(2):439–443, 1982.

O. Gossner. Comparison of information structures. *Games and Economic Behavior*, 30(1):44–63, 2000.

Eitan Greenshtein. Comparison of sequential experiments. *The Annals of Statistics*, 24(1):436–448, 1996.

E. Lehrer, D. Rosenberg, and E. Shmaya. Signaling and mediation in games with common interests. *Games and Economic Behavior*, 68(2):670–682, 2010.

Ehud Lehrer, Dinah Rosenberg, and Eran Shmaya. Garbling of signals and outcome equivalence. *Games and Economic Behavior*, 81:179–191, 2013.

Moshe Leshno and Yishay Spector. An elementary proof of Blackwell's theorem. *Mathematical Social Sciences*, 25(1):95–98, 1992.

Saunders Mac Lane. *Categories for the working mathematician*, volume 5. Springer Science & Business Media, 1978.

Marcin Pęski. Comparison of information structures in zero-sum games. *Games and Economic Behavior*, 62(2):732–735, 2008.