Measurability Is Not About Information *

Juan Dubra † and Federico Echenique ‡

February 7, 2002

Abstract

We present a simple example where the use of σ-algebras as a model of information leads to a paradoxical conclusion: a decision maker prefers less information to more. We then explain why the problem arises, and provide a characterization of the different models of information in the literature in terms of Blackwell's Theorem.

JEL Classification: C60, C70, G12

Keywords: Blackwell's Theorem, measurability, models of information, partitions, information-preserving σ-algebras.

*We are grateful to David Blackwell, Don Brown, Héctor Chade, David Levine, Jim Pitman, Maxwell Stinchcombe, Tarun Sabarwal, Bill Sudderth, Sergio Turner, Bill Zame, and seminar participants at UCLA and Yale for comments. We thank Larry Epstein, Peter Fishburn, Itzak Gilboa, Catherine Huafei Yan, Massimo Marinacci, Ernesto Mordecki, Malempati Rao, Srinivasa Varadhan and Peter Wakker for their kind attention to our questions. Part of this work was done while Juan Dubra visited Yale University, and Federico Echenique visited U.C. Berkeley

†Corresponding author. Universidad de Montevideo. `email: dubraj@um.edu.uy`

‡Departamento de Economía, Universidad Torcuato Di Tella, and Facultad de Ciencias Sociales, Universidad de la República, `email: fechenique@utdt.edu`.

§ 1 Introduction

We present a simple example where the use of σ-algebras as a model of information leads to a paradoxical conclusion: a decision maker prefers less information to more. We then explain why the problem arises, and provide a characterization of the different models of information in the literature in terms of Blackwell's Theorem.

§ 1.1 The Example

Let the state of the world be a real number between 0 and 1, so the set of possible states is $\Omega = [0, 1]$. Suppose that a decision maker can choose to either be perfectly informed, so that she gets to know the exact value of ω, or only be told if the true ω is smaller or larger than $1/2$. In the first case, the information can be modeled as the partition of all elements of Ω, $\tau = \{\{\omega\} : \omega \in \Omega\}$. In the second case, the information is the partition $\tau' = \{[0, 1/2) ; [1/2, 1]\}$.

The individual must first choose between τ and τ', and then decide to buy either a bond or a stock. The return on the stock is $S(\omega) = \omega$, the bond yields $3/8$ in every state of the world. Suppose in addition that the state of the world is chosen according to a uniform distribution on $[0, 1]$ and that the utility function is given by $u(\omega) = \omega$.

In contexts like this, it is common to model the informational content of τ and τ' by $\sigma(\tau)$ and $\sigma(\tau')$, the $\sigma-$algebras generated by the partitions. [1] We now show that doing this makes the decision maker choose τ' instead of τ. It is easy to check that for all $s \in \sigma(\tau)$, either s is countable, or s^c is countable. Then, as we will show in the appendix, $E(S|\sigma(\tau)) = 1/2$ a.s. (specifically, that any version of $E(S|\sigma(\tau))$ equals a.s. the constant function $1/2$). It is immediate that

$$E(S|\sigma(\tau'))(\omega) = \begin{cases} 1/4 & \text{if } \omega < 1/2 \\ 3/4 & \text{otherwise} \end{cases} .$$

In almost every state of the world $E(S|\sigma(\tau)) = 1/2$ is larger than the return to the bond $(3/8)$, so if the individual decides to observe information τ she will a.s. buy the stock. The expected utility of choosing τ is thus $1/2$. On the other hand if she chooses τ', she will buy the bond when $\omega < 1/2$ and the stock if $\omega \geq 1/2$, and will thus get a utility of $(3/8)(1/2) + (3/4)(1/2) = 9/16 > 1/2$.

Then, while τ is obviously more informative than τ', the decision maker strictly prefers τ' over τ. The reason is that the σ-algebra generated by τ, and used in forming $E(S|\sigma(\tau))$,

[1] Alternatively, one may use $\sigma-$algebras generated by signals (random variables) associated to the two information structures. As we discuss in section 3.1, the same phenomenon shows up.

is not informative at all: it is a collection of trivial sets, having either probability 0 or 1. The σ-algebra generated by τ, on the other hand, distinguishes $[0, 1/2)$ from $[1/2, 1]$.

§ 1.2 Our point.

The problem in our example is that finer partitions need not generate finer σ-algebras. We show below that σ-algebras do not preserve information because they are not closed under arbitrary unions. Heuristically, if the decision maker "knows" that some collection $\mathcal{E}$ of events is false, she should "know" that their union is false. Unless $\mathcal{E}$ is countable, however, the union of events in $\mathcal{E}$ need not be in the σ-algebra.

As is well-known, there are many technical problems with "'large" σ-algebras—and closedness under arbitrary unions would generally deliver large σ-algebras. We have no solution to offer, but we think it is important to document and explain these problems with the interpretation of σ-algebras.

§ 1.3 Notation and Definitions

A ***partition*** τ of a set Ω is a collection of pairwise disjoint subsets whose union is Ω; note that for each state of nature ω there is a unique element of τ that contains ω. A decision maker whose information is represented by τ is informed only that the element of τ that contains the true state of nature has occurred. In other words, the decision maker cannot distinguish between states that belong to the same element of τ. If τ, τ' are partitions, say that τ' is ***finer*** than τ, written $\tau' \geq \tau$, if for every $C \in \tau'$ there is B in τ such that $C \subseteq B$. For any collection $\mathcal{C}$ of subsets of Ω, the σ-algebra ***generated*** by $\mathcal{C}$, denoted $\sigma(\mathcal{C})$, is the smallest σ-algebra that contains $\mathcal{C}$.

We can define information structures to be signals $f : \Omega \to Y$, for some set Y of observable signal values. This way of modeling information is equivalent to using partitions: each signal generates a partition, and each partition can be interpreted as a signal. Define the partition P_f of Ω associated to f by

$$P_f = \left\{ f^{-1}(y) : y \in Y \right\}.$$

A pair $(\Omega, \mathcal{F})$, where $\mathcal{F}$ is a σ-algebra of subsets of Ω, is a measurable space. Let $(\Omega, \mathcal{F})$ and $(\Theta, \mathcal{G})$ be two measurable spaces. A function $f : \Omega \to \Theta$ is ***measurable*** if $f^{-1}(B) \in \mathcal{F}$ for all $B \in \mathcal{G}$. The σ-algebra generated by f, denoted $\sigma(f, \mathcal{G})$, is the smallest σ-algebra on Ω for which f is measurable. We say that a σ-algebra $\mathcal{B}$ on Y ***distinguishes*** f if for all $\omega \in \Omega$, $f(\omega) \in \mathcal{B}$.

For a given set of states of nature Ω and an arbitrary measurable set $(Y, \mathcal{B})$, a collection $\alpha = (m_\omega)_{\omega \in \Omega}$ of probability measures on $(Y, \mathcal{B})$ is an ***experiment***. If $\mathcal{B}, \mathcal{C}$ are σ-algebras on Y, W respectively, a stochastic transformation T is a function $Q(y, E)$ defined for all $y \in Y$ and $E \in \mathcal{C}$ which for fixed E is a measurable function of y and for fixed y is a probability measure on $\mathcal{C}$. For any probability measure m on $\mathcal{B}$, the function

$$M(E) = \int Q(y, E)\, dm(y)$$

is a probability measure on $\mathcal{C}$, denoted by Tm. If $\alpha = (m_\omega)_{\omega\in\Omega}$ and $\beta = (M_\omega)_{\omega\in\Omega}$ are two experiments, with m_ω, M_ω defined on $\mathcal{B}$ and $\mathcal{C}$ respectively, we shall say that α is ***sufficient*** for β, or β is a ***garbling*** of α, written $\alpha \succ \beta$ if there exists a stochastic transformation T such that $Tm_\omega = M_\omega$ for all ω. For $f : \Omega \to Y_f$ and $g : \Omega \to Y_g$, we will say that g is a garbling of f if $\beta = \left(\delta_{g(\omega)}\right)$ is a garbling of $\alpha = \left(\delta_{f(\omega)}\right)$. To understand this definition, notice that an experiment is just a function from Ω to the set of probability measures on some space $(Y_f, \mathcal{B})$. Then, a signal $f : \Omega \to Y_f$ can be identified with the experiment that associates with each ω, the lottery which is degenerate in $f(\omega)$.

§ 2 A Blackwell theorem.

Consider the following setup.

- Ω is the set of states of nature;
- Z is the set of consequences, Z has at least two elements;
- an ***act*** is a function $a : \Omega \to Z$, $A = Z^\Omega$ is the set of all acts;
- a ***decision maker*** is a preference relation $\succeq$ on A ($\succeq$ is a complete transitive binary relation on A).

The information structures available to a decision maker are signals $f : \Omega \to Y_f$ for some space Y_f. The decision maker is informed of the value taken by f and she must then choose a consequence in Z. An act $a : \Omega \to Z$ is f-feasible if $a(\omega) = a(\omega')$ whenever $f(\omega) = f(\omega')$.

A decision maker $\succeq$ ***prefers*** signal f to g if and only if for all g-feasible act a there exists a f-feasible act $\widetilde{a}$ such that $\widetilde{a} \succeq a$.

Theorem A. *Let $f : \Omega \to Y_f$ and $g : \Omega \to Y_g$. The following are equivalent.*

1. *Every decision maker prefers f to g;*
2. *P_f is finer than P_g;*

3. *There is* $h : Y_f \to Y_g$ *such that* $g = h \circ f$*;*

4. g *is a garbling of* f*;*

5. *the* σ*-algebra of arbitrary unions of elements in* P_f *is finer than the* σ*-algebra of arbitrary unions of elements in* P_g*.*

6. *for all* $\sigma-$*algebras* $\mathcal{B}$*,* $\mathcal{C}$ *on* Y_f, Y_g *that distinguish* f *and* g*, and are closed under arbitrary unions,* $\sigma(f, \mathcal{B})$ *is finer than* $\sigma(g, \mathcal{C})$*.*

Remarks

1. Blackwell's (1951) Theorem is the equivalence of 1 and 4, in the context of "noisy signals": his experiments are functions from Ω to the set of probability measures on some space X. By enlarging the state-space, his context can be embedded in ours, but the statement in Blackwell (1951) does not follow from our Theorem A applied to the enlarged state-space. On the other hand, Blackwell's theorem does not imply $1 \Leftrightarrow 4$ in the present context, as his theorem was for a finite state space (this is also true of the version in Blackwell (1953)). Therefore, neither theorem is more general.

2. We believe that the equivalence of 1, 2, 3, and 4 in Theorem A is known, but we are unaware of a statement or proof in print. In any case, we are interested in the equivalence of 1 with 5 and 6 as an explanation of our example in §1.1.

Theorem A explains the paradox in our example. Signal τ is more informative than τ', but the σ-algebra generated by τ is not finer than the σ-algebra generated by τ'. The root of the problem is very simple: if the decision maker knows that no ω with $\omega < 1/2$ has occurred, any model of information should prescribe that the decision maker knows that $[0, 1/2)$ has not occurred. But, if one models the decision maker's information as $\sigma(\tau)$, the decision maker can never "know" that the event $[0, 1/2)$ occurred or not, as $[0, 1/2) \notin \sigma(\tau)$. Knowledge is closed under arbitrary unions, but $\sigma-$algebras need not be.

§ 3 Discussion

1. A common alternative to modeling the information content of a signal $f : \Omega \to Y_f$, when Y_f is endowed with a σ-algebra $\mathcal{C}$, is $\sigma(f, \mathcal{C})$, the σ-algebra generated by f. By Theorem A this construction will preserve information if σ-algebras on target spaces are closed under arbitrary unions. If this is not the case, it is easy to generate examples where $\sigma(f, \mathcal{C})$ does not preserve information. In fact, by choosing the σ-algebras $\mathcal{C}$ and

$\mathcal{B}$ on target spaces appropriately, we can reproduce the example in the introduction with signals f, g, and conditional expectations $E(S|\sigma(f,\mathcal{C}))$ and $E(S|\sigma(g,\mathcal{B}))$. [2]

As an illustration, consider the following situation. Suppose $Y_f = Y_g = [0,1]$, and let $\mathcal{F}$ be any σ-algebra of subsets of $[0,1]$ such that $\{1\} \in \mathcal{F} \neq 2^{[0,1]}$. If $f : \Omega \to Y_f$ is the identity, and $g = \chi_E$ for some $E \in 2^{[0,1]} \backslash \mathcal{F}$, we obtain $\sigma(f,\mathcal{F}) = \mathcal{F}$, and $\sigma(g,\mathcal{F}) = \{\emptyset, E, E^c, [0,1]\}$. Thus $\sigma(f,\mathcal{F})$ and $\sigma(g,\mathcal{F})$ are not comparable, while f is more informative than g.

This example shows the same problems as those in the introduction may arise when $\mathcal{F}$ is chosen appropriately. The results of Blackwell (1956) show that under some regularity conditions on $(\Omega, \mathcal{B})$ and the $\sigma-$algebras on Y_f and Y_g, the $\sigma-$algebra generated by f is finer than that generated by g. This raises the issue of what $\sigma-$algebra should be used on the target space. Unfortunately, we do not have a preference-based theory for selecting among alternative $\sigma-$algebras on target spaces. [3] In addition, $\sigma(f,\mathcal{F})$ changes when $\mathcal{F}$ changes, and the information content of f does not.

One could interpret $\mathcal{C}$ as the events that the decision maker can observe, and use this interpretation in order to choose among different σ-algebras on the target spaces. But then $\mathcal{C}$ should also be closed under arbitrary unions, both to preserve information (Theorem A) and because "perception"—much like "knowledge"— should be closed under arbitrary unions.

2. When Ω is countable, Theorem A shows that σ-algebras preserve information, as any union of sets in Ω can at most be countable. Many models, though, require an uncountable number of states of nature. This is the case, for example, of Savage's model of decision under uncertainty (if we want subjective probabilities to be countably additive, as is usually the case in economics), or of games of incomplete information (Mertens and Zamir 1985, Brandenburger and Dekel 1993). In other models, an uncountable space is necessary to use calculus methods.

3. The information-preserving σ-algebras of arbitrary unions have two well-known disadvantages (see e.g. Dudley (1989)): they may be too large for some countably additive measures to be well-defined, and they have no clear links to the spaces' topological properties, like the Borel.

4. Stinchcombe (1990) proves that, in the spaces that Blackwell (1956) introduced, a countably generated σ-algebra $\mathcal{F}$ can be identified with the partition $\{\cap \{B : B \in \mathcal{F}, \omega \in B\} : \omega \in \Omega\}$ of its ***atoms***, and that all σ-algebras are "close" to a countably generated σ-algebra. [4]

[2] The signals $f, g : \Omega \to \Omega$, where f is the identity and g is the indicator function of the interval $[0, 1/2)$ work, when Ω is endowed with the countable-co-countable σ-algebra. Signal f is more informative than g, but the decision maker ends up preferring g.

[3] We thank Larry Epstein, Peter Fishburn, Itzak Gilboa, Massimo Marinacci and Peter Wakker for their feedback on this issue.

[4] We are grateful to Maxwell Stinchcombe for pointing this out.

In this particular sense, then, arbitrary σ-algebras possess an informational content.

5. Our example has some similarities with Example 4.10 in Billingsley (1995). Billingsley argues that the interpretation of σ-algebras as information is weak, using the following argument. First, he notices that the countable-co-countable σ-algebra $\mathcal{F}$ generates the finest partition of $[0,1]$ (two states belong to the same element of the partition if no $B \in \mathcal{F}$ distinguishes between them). In that sense, $\mathcal{F}$ contains "all the information". Second, he notices that for every $B \in \mathcal{F}$ and any Lebesgue measurable C, the Lebesgue probability of C conditional on B is just the Lebesgue probability of C. In that sense, $\mathcal{F}$ contains no information at all.

There is one important difference between Billingsley's example and our example: we argue that what makes the connection between information and σ-algebras weak is that finer partitions need not generate finer σ-algebras; Billingsley's argument is concerned with the partitions generated by σ-algebras. This distinction is relevant because different partitions generate different σ-algebras, but different σ-algebras may generate the same partition. In particular, Billingsley's example does not imply that Blackwell's theorem is false when information is modeled as σ-algebras—our example does.

5. The results in this paper are very simple, the question remains if they are not "known," or part of some oral tradition—we are confident they have not appeared in print. Yannelis (1991) claims that finer partitions generate finer σ-algebras. We do not wish to claim that any of his results are false, but we do believe Yannelis' paper is proof that our points are original.

§ 4 Appendix

Proof of $E(S|\sigma(P_f)) = 1/2$ in the example of the introduction. We shall show that any version of $E(S|\sigma(P_f))$ is a.s. equal to $1/2$. Let $j : \Omega \to \mathbf{R}$ be any version of $E(S|\sigma(P_f))$. We shall show that $j^{-1}(1/2, +\infty)$ and $j^{-1}(-\infty, 1/2)$ are countable sets; this suffices as countable sets have probability zero. We first show that, for arbitrary natural n, $j^{-1}\left[\frac{1}{2} + \frac{1}{n}, +\infty\right)$ is countable. Suppose it is uncountable, then $(j^{-1}\left[\frac{1}{2} + \frac{1}{n}, +\infty\right))^c$ is countable, because j is $\sigma(P_f)$-measurable and $\sigma(P_f)$ is the countable-co-countable σ-algebra. Then $j^{-1}\left[\frac{1}{2} + \frac{1}{n}, +\infty\right)$ has measure one and we obtain,

$$\frac{1}{2} = \int_{j^{-1}\left[\frac{1}{2}+\frac{1}{n},+\infty\right)} j(\omega)dP(\omega) \geq \int_{j^{-1}\left[\frac{1}{2}+\frac{1}{n},+\infty\right)} \left(\frac{1}{2} + \frac{1}{n}\right) dP = \frac{1}{2} + \frac{1}{n},$$

a contradiction. That $j^{-1}\left(\frac{1}{2}, +\infty\right)$ is countable follows, as

$$j^{-1}(1/2, +\infty) = \bigcup_n j^{-1}[1/2 + 1/n, +\infty).$$

Similarly, $j^{-1}(-\infty, 1/2)$ is countable. ■

Proof of Theorem (A). $(2 \Rightarrow 1)$. If $P_f \geq P_g$, all g-feasible acts are f-feasible. The result is immediate. To prove $(1 \Rightarrow 2)$, suppose, by way of contradiction, that P_f is not finer than P_g. Then there exist ω and ω' such that $f(\omega) = f(\omega')$ but $g(\omega) \neq g(\omega')$. Define, for $x \neq y$ the following act

$$a(\omega'') = \begin{cases} x & \text{if } g(\omega'') = g(\omega) \\ y & \text{otherwise} \end{cases}$$

Note that a is g-feasible but not f-feasible. Let $\succeq$ be a decision maker such that $a \succ \widetilde{a}$ for all $\tilde{a} \in A$. There is no f-feasible $\tilde{a} \in A$ such that $\tilde{a} \succeq a$, so $\succeq$ does not prefer f to g.

$(2 \Rightarrow 3)$ Suppose $P_f \geq P_g$. For each $z \in g(\Omega)$ let $Y_f^z = f(g^{-1}(z))$ and let $\widetilde{z}$ be any element in Z. It is easy to check that for

$$h(y) = \begin{cases} z & \text{if } y \in Y_f^z \text{ for some } z \text{ in } g(\Omega) \\ \widetilde{z} & \text{otherwise} \end{cases}$$

we have $h \circ f = g$.

$(3 \Rightarrow 2)$ If it is not the case that $P_f \geq P_g$, there exist ω and ω' such that ω and ω' are in the same element of P_f but not of P_g. Then, for all $h : Y_f \to Y_g$,

$$h(f(\omega)) = h(f(\omega')) \text{ but } g(\omega) \neq g(\omega')$$

so $h \circ f \neq g$.

$(2 \Rightarrow 4)$. Notice that $Q(y, \cdot) = \delta_{gf^{-1}(y)}(\cdot)$ satisfies

$$\begin{aligned} T\delta_{f(\omega)}(E) &= \int Q(y, E)\, d\delta_{f(\omega)}(y) = Q(f(\omega), E) \\ &= \delta_{gf^{-1}(f(\omega))}(E) = \delta_{g(\omega)}(E) \end{aligned}$$

as was sought.

$(4 \Rightarrow 2)$. By hypothesis, $\delta_{g(\omega)}(E) = \int Q(y, E) d\delta_{f(\omega)}(y)$ for some Q. Then, $f(\omega) = f(\omega')$ implies that $\delta_{g(\omega)}(E) = \delta_{g(\omega')}(E)$. So $g(\omega) = g(\omega')$. ■

$(2 \Rightarrow 5)$ If $P_f \geq P_g$, for every collection of sets $\{C_i\}$ in P_g, there exists a collection $\{B_i\}$ in P_f such that $\bigcup_i C_i = \bigcup_i B_i$. Thus, letting $\alpha(\tau)$ stand for the σ-algebra of arbitrary unions of elements in τ, $\alpha(P_g) \subseteq \alpha(P_f)$.

$(5 \Rightarrow 2)$ Suppose that it is not the case that $P_f \geq P_g$. This means that there exists a set C in P_g such that for every collection $\{B_i\}$ in P_f, $C \neq \bigcup_i B_i$. Thus, $C \in \alpha(P_g)$, but $C \notin \alpha(P_f)$, a contradiction.

$(2 \Rightarrow 6)$ Suppose $P_f \geq P_g$, and that $c \in \sigma(g, \mathcal{C})$. This means that there exists $C \in \mathcal{C}$ such that $c = g^{-1}(C)$. Since $P_f \geq P_g$,

$$c = \bigcup_{i \in I} p_i$$

for some index set I and $p_i \in P_f$ for all i. Since $\mathcal{B}$ distinguishes f, $f(p_i) \in \mathcal{B}$ for all i, and since $\mathcal{B}$ is closed under arbitrary unions, $\bigcup_{i \in I} f(p_i) \in \mathcal{B}$. We then obtain that

$$c = f^{-1}\left(\bigcup_{i \in I} f(p_i)\right) \in \sigma(f, \mathcal{B}).$$

$(6 \Leftarrow 2)$ It is easy to check that since $\mathcal{B}$ ($\mathcal{C}$ resp) distinguishes f (g resp.), $\sigma(f, \mathcal{B})$ ($\sigma(g, \mathcal{C})$ resp.) is the collection of arbitrary unions of elements of P_f (P_g resp.). Therefore, for all $c \in P_g$ we have $c \in \sigma(g, \mathcal{C})$, and by hypothesis, $c \in \sigma(f, \mathcal{B})$. But then, there must exist there exist a collection $\{p_i\}_{i \in I}$ such that $p_i \in P_f$ for all i, and $c = \bigcup_{i \in I} p_i$, so $P_f \geq P_g$. ■

References

BILLINGSLEY, P. (1995): *Probability and Measure.* J. Wiley & Sons.

BLACKWELL, D. (1951): "Comparison of Experiments," in *Proceedings of the second Berkeley symposium on mathematical statistics and probability*, ed. by J. Neyman, pp. 93–102. University of California Press, Berkeley and Los Angeles.

——— (1953): "Equivalent Comparisons of Experiments," *Annals of Mathematical Statistics*, 24(2), 93–102.

——— (1956): "On a Class of Probability Spaces," in *Proceedings of the third Berkeley symposium on mathematical statistics and probability*, ed. by J. Neyman, pp. 1–6. University of California Press, Berkeley and Los Angeles.

BRANDENBURGER, A., AND E. DEKEL (1993): "Hierarchies of Beliefs and Common Knowledge," *Journal of Economic Theory*, 59(1), 189–98.

DUDLEY, R. M. (1989): *Real Analysis and Probability.* Chapman & Hall.

MERTENS, J.-F., AND S. ZAMIR (1985): "Formulation of Bayesian Analysis for Games of Incomplete Information," *International Journal of Game Theory*, 14, 1–29.

Stinchcombe, M. B. (1990): "Bayesian Information Topologies," *Journal of Mathematical Economics*, 19, 233–253.

Yannelis, N. C. (1991): "The Core of an Economy with Differential Information," *Economic Theory*, 1(2), 183–97.