

Dilation, Disintegrations, and Delayed Decisions

Arthur Paul Pedersen

Center for Adaptive Rationality
Max Planck Institute for Human Development
Lentzeallee 94, 14195 Berlin
pedersen@mpib-berlin.mpg.de

Gregory Wheeler

Munich Center for Mathematical Philosophy
Ludwig Maximilians University
Geschwister-Scholl-Platz 1, 80539 Munich
gregory.wheeler@lrz.uni-muenchen.de

Abstract

Both dilation and non-conglomerability have been alleged to conflict with a fundamental principle of Bayesian methodology that we call *Good's Principle*: one should always delay making a terminal decision between alternative courses of action if given the opportunity to first learn, at zero cost, the outcome of an experiment relevant to the decision. In particular, both dilation and non-conglomerability have been alleged to permit or even mandate choosing to make a terminal decision in deliberate ignorance of relevant, cost-free information. Although dilation and non-conglomerability share some similarities, some authors maintain that there are important differences between the two that warrant endorsing different normative positions regarding dilation and non-conglomerability. This article reassesses the grounds for treating dilation and non-conglomerability differently. Our analysis exploits a new and general characterization result for dilation to draw a closer connection between dilation and non-conglomerability.

1 Introduction

Good's Principle is considered by I. J. Good [8], among others before him [24, 19, 25], to be a fundamental principle of rational decision making. Good's Principle recommends to delay making a terminal decision between alternative courses of action if the opportunity arises to learn, at no cost, the outcome of an experiment relevant to the decision.

Dilation [34, 31, 21] occurs when an interval estimate of an event E is properly included in the interval estimate of E conditional on the occurrence of every event of a measurable partition \mathcal{B} . In such circumstances merely running the experiment to determine the value of \mathcal{B} , whatever the outcome, suffices to render your initial estimate of E less precise. Should you update your estimate of E to the less precise estimate? Should you refuse a free offer to learn the outcome of such an

experiment? Is it rational for you to pay someone to *not* tell you?

A probability function p is *non-conglomerable* [4, 5] for an event E in a measurable partition \mathcal{B} if the marginal probability of E fails to be included in the closed interval determined by the infimum and supremum of the set of conditional probabilities of E given each cell of \mathcal{B} . When \mathcal{B} is denumerable, any probability function is non-conglomerable for E in \mathcal{B} only if it fails to be countably additive [4, 5, 26]. In such circumstances merely running the experiment to determine the value of \mathcal{B} , whatever the outcome, suffices to uniformly increase (or decrease) your initial estimate of E . Is your initial estimate of E coherent? Is it rational to forgo the opportunity to learn the experimental outcome of \mathcal{B} ?

Even though both dilation [9, 6] and non-conglomerability [26] have been alleged to conflict with Good's Principle, there is a tradition within the imprecise probability community to treat each differently. Walley, for example, argues that conglomerability is a requirement of rationality in the course of extending coherent lower previsions to conditional lower previsions. More recently, Zaffalon and Miranda argue that conglomerability is a requirement of rationality when an agent's future commitments and current conditional beliefs are established together [36]. Either way, instances of non-conglomerability generate violations of salient dominance principles and allow for the devaluation of cost-free information and thus violations of Good's Principle [14]. Even so, instances of dilation do not preclude violations of salient dominance principles and of Good's Principle – see §5, below – but dilation is viewed as a reasonable, even if surprising, feature of conditional lower previsions [34, §6.4.3]. For Seidenfeld et al. [26], non-conglomerability raises a challenge for those who concede that sometimes rationality permits credal states to be representable by numerically precise probabilities failing to be countable additive. More specifically, Seidenfeld et al. observe that every

instance of non-conglomerability can be transformed into a violation of admissibility, a dominance principle at the heart of the Bayesian enterprise (e.g., Wald, de Finetti, Savage), and that expected utility maximization in such cases admits the devaluation of cost-free information. However, for some decision rules proposed for imprecise probabilities, such as Γ -maximin, dilation also invites a devaluation of cost-free information. But in this case Seidenfeld recommends to reject the decision rule rather than dilation [29].

It is true that while failures of conglomerability can only occur only in infinite partitions, dilation can occur with respect to finite partitions. This observation alone, of course, fails to provide an adequate explanation for adopting a view that treats dilation and non-conglomerability differently with respect to similar problems. In this paper we challenge the practice of treating dilation and non-conglomerability differently. Our analysis appeals to a new and general characterization result for dilation to draw a closer connection between dilation and non-conglomerability

The structure of the paper is as follows. In §2 we review dilation and present our general characterization result purely in terms of distance from independence. Then, in §3 we review the conglomerability principle and rehearse a standard example of non-conglomerability. In §4, we discuss Good's Principle in more detail and introduce a general framework within which to express Good's Principle, as it is commonly understood, in terms of subjective expected utility. Then, in §5 we discuss various violations of Good's Principle, with special attention to two examples in particular, one involving dilation and the other involving non-conglomerability. In particular, we argue that the normative standing of Good's Principle in the dilation case depends on particular features of the uncertainty model and the decision rules used, both of which depend ultimately on the decision maker's beliefs, values and goals. We then turn to an example involving non-conglomerability to argue that such examples should be treated in the same fashion, that is, that the normative standing of conglomerability likewise depends on the features of the uncertainty model and decision rules the decision maker uses.

2 Dilation

A *lower probability space* is a quadruple $(\Omega, \mathcal{A}, \mathbb{P}, \underline{P})$ such that Ω is a set of states, \mathcal{A} is an algebra over Ω , \mathbb{P} is a nonempty set of probability functions on \mathcal{A} , and \underline{P} is a *lower probability function* on \mathcal{A} with respect to \mathbb{P} —that is, $\underline{P}(E) = \inf\{p(E) : p \in \mathbb{P}\}$ for each $E \in \mathcal{A}$. The value $\underline{P}(E)$ is called the *lower probability* of E . The *upper probability function* \bar{P} is

then defined in the usual manner by stipulating that $\bar{P}(E) = 1 - \underline{P}(E^c)$ for each $E \in \mathcal{A}$; the value $\bar{P}(E)$ is called the *upper probability* of E . If $\underline{P}(H) > 0$, then conditional lower and upper probabilities are defined as $\underline{P}(E | H) = \inf\{p(E | H) : p \in \mathbb{P}\}$ and $\bar{P}(E | H) = \sup\{p(E | H) : p \in \mathbb{P}\}$, respectively. In the following, we call a collection of events \mathcal{B} from \mathcal{A} a *positive measurable partition* (of Ω) if \mathcal{B} is a partition of Ω such that $\underline{P}(H) > 0$ for each $H \in \mathcal{B}$.

Let \mathcal{B} be a positive measurable partition of Ω . We say that \mathcal{B} *dilates* E just in case for each $H \in \mathcal{B}$:

$$\underline{P}(E | H) < \underline{P}(E) \leq \bar{P}(E) < \bar{P}(E | H).^1$$

In other words, \mathcal{B} dilates E just in case the closed interval $[\underline{P}(E), \bar{P}(E)]$ is contained in the open interval $(\underline{P}(E | H), \bar{P}(E | H))$ for each $H \in \mathcal{B}$.

What is remarkable about dilation is the specter of turning a more precise estimate of E into a less precise estimate, no matter what event from the partition occurs.

Next, in §2.1, we rehearse an example from [28] involving a maximally uncertain event, G , a flip of a fair coin (whose outcomes form a partition, \mathcal{B}) and a pivotal quantity, E , defined in terms of G and the outcome of the coin toss. Then, in §2.3, we provide a simple characterization of dilation in terms of distance from stochastic independence, followed by a short discussion of the result.

2.1 Example of Dilation

Suppose G is a highly uncertain event, one with upper probability close to 1, $\bar{P}(G) = .9$, and lower probability close to 0, $\underline{P}(G) = .1$. So,

$$\bar{P}(G) - \underline{P}(G) = 0.8. \quad (1)$$

Suppose now that $\mathcal{B} = \{H, H^c\}$ is a partition representing the outcomes of a fairly tossed coin,

$$\underline{P}(H) = \bar{P}(H) = \frac{1}{2} = \underline{P}(H^c) = \bar{P}(H^c). \quad (2)$$

In addition to being positively measurable, suppose the outcomes of the toss are *stochastically independent* of our maximally uncertain event. In particular, the event of the coin landing heads, H , is stochastically independent of G occurring; hence, for each $p \in \mathbb{P}$,

$$p(G \cap H) = p(G)p(H) = \frac{p(G)}{2}. \quad (3)$$

Next let E be the event of either G and H both occurring or both failing to occur, namely $E := (G \cap$

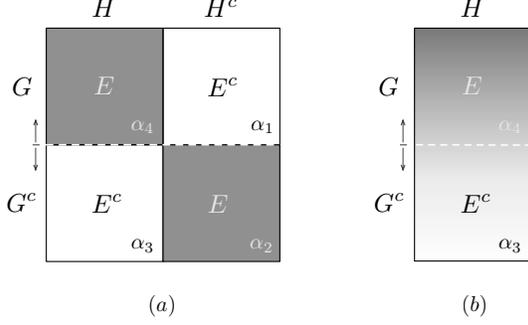


Figure 1: (a) 2x2 Table for an uncertain event (row) and a fair coin randomizer (column); (b) The event of learning that the outcome of the coin toss is ‘heads’.

$H) \cup (G^c \cap H^c)$. The probability of E is determinate: $p(E) = \frac{1}{2}$. Similarly, for each $p \in \mathbb{P}$, $p(E^c) = 1/2$.

These conditions are represented by the two-by-two table in Figure 1(a). Here the *columns* H and H^c represent the two possible outcomes of the fair coin toss; the *rows* G and G^c represent the two outcomes of our maximally uncertain event; the *diagonals* E and E^c describe the two events defined in terms of the possible outcomes of row and column: E is the “matching” event of either H and E both obtaining or neither obtaining, and E^c is the “unmatched” event of one but not the other obtaining.

Observe that E is dilated by $\mathcal{B} = \{H, H^c\}$: although the initial estimate of E is precisely one-half, learning the outcome of the coin toss, whether heads or tails, dilates the probability estimate of E to $[.1, .9]$.

Proof - We show that $0.1 = \underline{P}(E | H) < \underline{P}(E) = 1/2$.

$$\begin{aligned} \underline{P}(E | H) &= \inf \{ p(E | H) : p \in \mathbb{P} \} \\ &= \inf \left\{ \frac{p([(G \cap H) \cup (G^c \cap H^c)] \cap H)}{p(H)} : p \in \mathbb{P} \right\} \\ &= \inf \left\{ \frac{p(G \cap H)}{p(H)} : p \in \mathbb{P} \right\} \\ &= \inf \left\{ \frac{p(G)p(H)}{p(H)} : p \in \mathbb{P} \right\} \\ &= 0.1 \end{aligned}$$

A similar argument establishes $0.9 = \overline{P}(E | H) > 1/2$, and the argument holds if instead the coin lands tails, i.e., $\underline{P}(E | H^c) = 0.1$ and $\overline{P}(E | H^c) = 0.9$. Thus, E is dilated by the coin toss, $\mathcal{B} = \{H, H^c\}$. \diamond

The specific case where the coin lands H is illustrated in Figure 1(b). Here conditioning on H reduces the probability that E obtains to the probability that G obtains, which is highly uncertain.

¹While our terminology agrees with that of [11, p. 252], it differs from that of [31, p. 1141] and [12, p. 412], who call dilation in our sense *strict dilation*.

2.2 Measuring Distance from Independence

Given a single probability function p on \mathcal{A} and events E and H with positive probability, the degree to which two events E and H diverge from stochastic independence, if they diverge at all, may be characterized by a simple measure of distance from stochastic independence:

$$S_p(E, H) := \frac{p(E \cap H)}{p(E)p(H)}.$$

The measure S_p is simply the covariance of E and H , $Cov(E, H) = p(E \cap H) - p(E)p(H)$, put in ratio form. Therefore, $S_p(E, H) = 1$ just in case E and H are stochastically independent; $S_p(E, H) > 1$ when E and H are positively correlated; and $S_p(E, H) < 1$ when E and F are negatively correlated. The measure S_p naturally extends to a set of probability functions \mathbb{P} as follows:

$$\begin{aligned} S_{\mathbb{P}}^+(E, H) &:= \{p \in \mathbb{P} : S_p(E, H) > 1\}; \\ S_{\mathbb{P}}^-(E, H) &:= \{p \in \mathbb{P} : S_p(E, H) < 1\}; \\ I_{\mathbb{P}}(E, H) &:= \{p \in \mathbb{P} : S_p(E, H) = 1\}. \end{aligned}$$

The set of probability functions $I_{\mathbb{P}}(E, H)$ from \mathbb{P} with E and H stochastically independent is called *the surface of independence* for E and H with respect to \mathbb{P} . In the remainder subscripts will be dropped when there is no danger of confusion.

2.3 A Simple Characterization of Dilation

In this section, we present simple necessary and sufficient conditions for dilation formulated in terms of deviation from stochastic independence, which improves upon previous results in [21]. We illustrate an immediate application of such a characterization with measures of dilation. To begin, we introduce the notion of a neighborhood.

Given a lower probability space $(\Omega, \mathcal{A}, \mathbb{P}, \underline{P})$, events $E, H \in \mathcal{A}$ with $\underline{P}(H) > 0$, and $\epsilon > 0$ define:

$$\begin{aligned} \underline{\mathbb{P}}(E|H, \epsilon) &:= \{p \in \mathbb{P} : |p(E|H) - \underline{P}(E|H)| < \epsilon\}; \\ \overline{\mathbb{P}}(E|H, \epsilon) &:= \{p \in \mathbb{P} : |p(E|H) - \overline{P}(E|H)| < \epsilon\}. \end{aligned}$$

We call the sets $\underline{\mathbb{P}}(E|H, \epsilon)$ and $\overline{\mathbb{P}}(E|H, \epsilon)$ lower and upper *neighborhoods* of E conditional on H with radius ϵ , respectively. Thus, a probability function p is an element of $\underline{\mathbb{P}}(E|H, \epsilon)$ if $p(E|H)$ is within ϵ of $\underline{P}(E|H)$, and similarly for an upper neighborhood.

For the sake of readability in what follows, given a nonempty set of probabilities \mathbb{P} , let \mathbb{P}_* denote $\overline{\text{co}}(\mathbb{P})$, the weak*-closed convex hull of \mathbb{P} . Thus, $\underline{\mathbb{P}}_*(E|H, \epsilon) =$

$\overline{\text{co}}(\mathbb{P})(E|F, \epsilon)$ and $\overline{\mathbb{P}}_*(E|F, \epsilon) = \overline{\overline{\text{co}}(\mathbb{P})}(E|F, \epsilon)$. Similarly, let $S_*^+(E, F)$ and $S_*^-(E, F)$ be defined by:

$$S_*^+(E, F) := \{p \in \overline{\text{co}}(\mathbb{P}) : S_p(E, F) > 1\}$$

$$S_*^-(E, F) := \{p \in \overline{\text{co}}(\mathbb{P}) : S_p(E, F) < 1\}.$$

Given a nonempty set I , we let \mathbb{R}_+^I denote the set of elements $(r_i)_{i \in I}$ of \mathbb{R}^I such that $r_i > 0$ for each $i \in I$. We now state a result characterizing dilation and then report an immediate corollary.

Theorem 1 Let $(\Omega, \mathcal{A}, \mathbb{P}, \underline{\mathbb{P}})$ be a lower probability space, let $\mathcal{B} = \{H_i : i \in I\}$ be a positive measurable partition, and let $E \in \mathcal{A}$. Then the following are equivalent:

- (i) \mathcal{B} dilates E ;
- (ii) There is $(\epsilon_i)_{i \in I} \in \mathbb{R}_+^I$ such that for every $i \in I$:

$$\underline{\mathbb{P}}_*(E|H_i, \epsilon_i) \subseteq S_*^-(E, H_i) \text{ and}$$

$$\overline{\mathbb{P}}_*(E|H_i, \epsilon_i) \subseteq S_*^+(E, H_i);$$

- (iii) There is $(\epsilon_i)_{i \in I} \in \mathbb{R}_+^I$ such that for every $i \in I$:

$$\underline{\mathbb{P}}(E|H_i, \epsilon_i) \subseteq S^-(E, H_i) \text{ and}$$

$$\overline{\mathbb{P}}(E|H_i, \epsilon_i) \subseteq S^+(E, H_i),$$

where for each $i \in I$, $\epsilon_i \leq \min(\underline{\epsilon}_i, \bar{\epsilon}_i)$ and $\underline{\epsilon}_i$ is the unique minimum of $|p(E|H_i) - \underline{\mathbb{P}}(E|H_i)|$ attained on $C_i^+ =_{df} \{p \in \mathbb{P}^* : S_p(E, H_i) \geq 1\}$, and $\bar{\epsilon}_i$ is the unique minimum of $|p(E|H_i) - \overline{\mathbb{P}}(E|H_i)|$ attained on $C_i^- =_{df} \{p \in \mathbb{P}^* : S_p(E, H_i) \leq 1\}$. \diamond

Theorem 1 implies that a positive measurable partition \mathcal{B} dilates an event E just in case for each partition cell H , there are upper and lower neighborhoods of E conditional on H such that the lower neighborhood of E conditional on H lies entirely within the subset of the set of probabilities in question for which E and H are negatively correlated, while the upper neighborhood of E given H lies entirely within the subset of the set of probabilities in question for which E and H are positively correlated. We remark that Theorem 1 holds for *arbitrary* nonempty sets of probabilities.

When \mathcal{B} is a finite positive measurable partition, the preceding theorem may be simplified.

Corollary 1 Let $(\Omega, \mathcal{A}, \mathbb{P}, \underline{\mathbb{P}})$ be a lower probability space, let $\mathcal{B} = (H_i)_{i=1}^n$ be a finite positive measurable partition, and let $E \in \mathcal{A}$. Then the following are equivalent:

- (i) \mathcal{B} dilates E ;

- (ii) There is $\epsilon > 0$ such that for each $i = 1, \dots, n$:

$$\underline{\mathbb{P}}_*(E|H_i, \epsilon) \subseteq S_*^-(E, H_i) \text{ and}$$

$$\overline{\mathbb{P}}_*(E|H_i, \epsilon) \subseteq S_*^+(E, H_i);$$

- (iii) There is $\epsilon > 0$ such that for each $i = 1, \dots, n$:

$$\underline{\mathbb{P}}(E|H_i, \epsilon) \subseteq S^-(E, H_i) \text{ and}$$

$$\overline{\mathbb{P}}(E|H_i, \epsilon) \subseteq S^+(E, H_i),$$

where $\epsilon \leq \min(\underline{\epsilon}_i, \bar{\epsilon}_i : i = 1, \dots, n)$ and $\underline{\epsilon}_i$ is the unique minimum of $|p(E|H_i) - \underline{\mathbb{P}}(E|H_i)|$ attained on $C_i^+ =_{df} \{p \in \mathbb{P}^* : S_p(E, H_i) \geq 1\}$, and $\bar{\epsilon}_i$ is the unique minimum of $|p(E|H_i) - \overline{\mathbb{P}}(E|H_i)|$ attained on $C_i^- =_{df} \{p \in \mathbb{P}^* : S_p(E, H_i) \leq 1\}$. \diamond

Thus, when the positive measurable partition \mathcal{B} is finite, the radii ϵ_i of Theorem 1 may be replaced by a *single* positive radius ϵ . The preceding corollary also improves upon a similar result in [21].

Discussion. Theorem 1 and Corollary 1 should hardly be surprising. The correlation properties that entail dilation are rather straightforward consequences of the definition. Moreover, these correlation properties entail that each dilating partition cell and dilated event live on the surface of independence under *some* probability function from the closed convex hull of the set of probabilities in question. Although straightforward, Theorem 1 shows that by looking *beyond* the upper and lower supporting hyperplanes $\overline{\mathbb{P}}_*(E|H)$ and $\underline{\mathbb{P}}_*(E|H)$ to the upper and lower supporting *neighborhoods* $\overline{\mathbb{P}}_*(E|H, \epsilon)$ and $\underline{\mathbb{P}}_*(E|H, \epsilon)$, it becomes possible to characterize dilation completely in terms of positive and negative correlation, achieving a longstanding goal. The results also show that dilation, properly understood, is a property of the *convex closure* of a set of probabilities.

One may see the generality of Theorem 1 by comparing it to an earlier result in [35, Result 1]. Observe that (1) Theorem 1 applies to *arbitrary* positive measurable partitions, whereas [35, Result 1] applies only to binary partitions; (2) Theorem 1 applies to *arbitrary* sets of probabilities, whereas Result 1 just applies to weak*-closed convex sets of probabilities; and (3) Theorem 1 presents characterizing conditions—property (ii) and property (iii)—formulated in terms of upper and lower neighborhoods, whereas Result 1 gives a characterizing condition formulated in terms of a patchwork of infimums and supremums—a point we discuss further in [21, §4]. Of course, Theorem 1, given its generality, entails that the characterizing condition of Result 1 in [35] is logically equivalent to property (ii)—or property (iii)—of Theorem 1 in the very special case for binary partitions and weak*-closed convex sets of

probabilities. Yet, in our judgment, the characterizing condition of Result 1, even with its narrow scope, is periphrastic. The upshot is that Theorem 1, in spite of its wide scope, delivers characterizing conditions which succinctly express the *wherefore* of dilation.

Last, returning to the simple heuristic example of dilation we presented in §2, we remark that a straightforward calculation of the relevant radii $\underline{\epsilon}_1, \bar{\epsilon}_1, \underline{\epsilon}_2, \bar{\epsilon}_2$ corresponding to H and H^c , respectively, yields $\frac{2}{5}$.

2.4 Proper and Improper Dilation

It is well known that the familiar univocal notion of probabilistic independence splinters into a plurality of logically distinct independence concepts [34, 2]. Thus, if a decision modeler knows that one event is epistemically independent of another – that is, that each event is epistemically irrelevant to the other – then he knows that observing the outcome of one event does not change the estimate in the other, even though the two events may fail to be stochastically independent, and thus may admit dilation. In other words, our characterization results hold for a variety of extensions—including unknown interaction, irrelevant natural extensions, and independent natural extensions [2]—without discriminating between models which correctly or incorrectly encode knowledge of either epistemic irrelevance or epistemic independence. However, our proposal is that a correctly parameterized extension *can* provide principled grounds for avoiding the loss of precision by dilation that may otherwise come from updating. So, even if the conditions for Theorem 1 hold, there may be enough knowledge about the relationship between the two events in question to support a parameterization that defuses the diluting effect that dilation has from updating. We therefore distinguish between two kinds of dilation phenomenon: *proper dilation*, which occurs within a model that correctly parameterizes the set of distributions to reflect what is known about how the events are interrelated, if anything is known at all, and *improper dilation*, which occurs within a model whose parameterization does not correctly represent what is known about how the events interact.

3 Non-Conglomerability

Given a real-valued finitely additive probability function p on an (σ) -algebra \mathcal{A} over a set of states Ω , a positive measurable partition \mathcal{B} of Ω , and an event E of \mathcal{A} , we say that p is *conglomerable* for E in \mathcal{B} if

$$\inf \{p(E|H) : H \in \mathcal{B}\} \leq p(E) \leq \sup \{p(E|H) : H \in \mathcal{B}\}$$

Otherwise we say that p is *non-conglomerable* for E in

\mathcal{B} . So p is non-conglomerable for E in \mathcal{B} just in case $p(E)$ fails to lie in the closed interval $[\inf \{p(E|H) : H \in \mathcal{B}\}, \sup \{p(E|H) : H \in \mathcal{B}\}]$.

Of course, every probability function is conglomerable for all events and finite \mathcal{B} . Cases of non-conglomerability only arise for infinite \mathcal{B} . It is well-known that any probability function with an infinite range is conglomerable for each event E and denumerable \mathcal{B} just in case it is countably additive. In addition, any such probability function is non-conglomerable for some event E and denumerable \mathcal{B} just in case it fails to be *disintegrable* for E in \mathcal{B} —that is, if fails to satisfy the law of total probability for E in \mathcal{B} . These concepts and results can be extended to bounded random quantities [5] and unbounded random quantities [27]. Further, it should be noted that some probability functions that fail to be countably additive may nonetheless be conglomerable in arbitrary positive measurable partitions. Moreover, in some cases, a nontrivial convex combination of probability functions, each of which fails to be conglomerable in a positive measurable partition, may very well be conglomerable in the partition. Indeed, a nontrivial convex combination of probability functions, each of which *is* conglomerable in a positive measurable partition, may very well fail to be conglomerable in the partition. To gain control over these cases, authors investigated conglomerability within the setting of primitive conditional probability, which accommodates conditioning events with zero probability [1, 26, 30]. Next we give an example of non-conglomerability for a denumerable partition.

3.1 Example of Non-Conglomerability

Following [5],² let \mathcal{A} be the collection of all subsets of $\Omega = \{0, 1\} \times \mathbb{N}_{>0}$, let $E = \{(1, n) : n \in \mathbb{N}_{>0}\}$, and let $\mathcal{B} = \{H_n : n \in \mathbb{N}_{>0}\}$, where $H_n = \{(0, n), (1, n)\}$ for each $n \in \mathbb{N}_{>0}$. Let p be a finitely-additive probability function on \mathcal{A} such that:

- (i) $p(E) = \frac{1}{2}$;
- (ii) $p(E \cap H_n) = \frac{1}{2^{n+1}}$ for each $n \in \mathbb{N}_{>0}$; and
- (iii) $p(E^c \cap H_n) = 0$.

Then $p(E) < \inf \{p(E|H_n) : n \in \mathbb{N}_{>0}\} = 1$, so p is non-conglomerable for E in the denumerable partition \mathcal{B} .

²This example seems to have entered the literature in [3, p. 205], although de Finetti there reports that Lester Dubins presented the example in a letter to L.J. Savage.

4 Good’s Principle and Expected Utility

In *Foundations of Statistics*, Savage considers the difference between a **basic** decision problem, in which an agent is to choose to perform one action from among several he judges to be available for choice, and a **derived** decision problem, in which the agent is to choose from the same basic actions, but only after considering the associated conditional expected utilities for the basic action given each possible outcome of some experiment. “It is almost obvious,” Savage remarks, “that the value of a derived problem cannot be less than, and typically is greater, than the value of the basic problem from which it is derived” [25, §6.2]. Savage thereupon formulates and proves what has become a fundamental principle of Bayesian methodology [25, Chapter 7]. Although Ramsey [24] aired the idea of this result in unpublished work and many others have reaffirmed it following Savage’s seminal work (e.g., [22], [19]), Good famously defended the principle in a short article published in the 1960s [8] – and there has been a rich discussion ever since [7, 33, 20, 28, 9, 32, 13]. Following Stigler’s law of eponymy, let us briefly explain the basic idea of *Good’s Principle*.

4.1 Formalizing Good’s Principle

Here is the set up. Suppose that at some time t_1 you are to face a choice among several courses of action a_1, \dots, a_n . Prior to this choice, however, you face a decision at some time t_0 before t_1 between (i) choosing from among several courses of action a_1, \dots, a_n at time t_1 or (ii) choosing from among the same courses of action a_1, \dots, a_n at some later time t_2 after you have observed, at no cost, the outcome of an experiment \mathfrak{E} .

According to Good’s Principle, Bayesian standards prohibit you from rejecting the opportunity to choose from among a_1, \dots, a_n at t_2 after observing the outcome of the experiment \mathfrak{E} . In addition, if the experimental outcome might affect your choice from among the courses of action, then Bayesian standards prohibit you from deciding to choose from among a_1, \dots, a_n at t_1 . In short, to be a Good Bayesian, take Good’s advice: accessible cost-free information relevant to a decision should never be ignored.

As a piece of Bayesian legislation, Good’s Principle is expressed in the legalese for codifying norms of classical subjective expected utility theory. In order to express Good’s Principle in the language of subjective expected utility, we first introduce the formal framework we shall use in our discussion. This framework is sufficiently expressive for our purposes and will enable us to carry out our discussion while remaining neutral over further

controversial matters unrelated to our concerns.

Let Ω be a set of states corresponding to a collection of hypotheses which are individually consistent, mutually exclusive, and collectively exhaustive relative to your state of certainty at time t_0 . A set of actions A is said to be a *decision problem* for you at time t if it consists of all actions you judge to be available for you to choose. Suppose that for each action a from A and each state ω in Ω , you have identified a unique consequence $\sigma(a, \omega)$ to be relevant for evaluating the action’s success in promoting the goals and values you endorse. So, you recognize that if you augment your state of certainty with the hypothetical supposition that you have implemented action a and state ω obtains, your transformed state of certainty commits you to being certain that consequence $\sigma(a, \omega)$ prevails.

We presume that you endorse a standard for decision making that commits you to identifying a nonempty subset $\mathfrak{c}(A)$ of your feasible actions A you judge to be *admissible*, or acceptable for choice, given your beliefs, values and goals.

Turn now to Good’s Principle illustrated in Figure 2. Suppose that you endorse subjective expected utility maximization as your standard. To sidestep some technical issues, suppose in particular that your judgments of admissibility can be represented in terms of subjective expected utility maximization with respect to a real-valued expectation $\mathbb{E}_p[\cdot]$ agreeing with a real-valued probability function p defined on a Boolean algebra over the set of states and a real-valued utility function u defined over the set of consequences.³

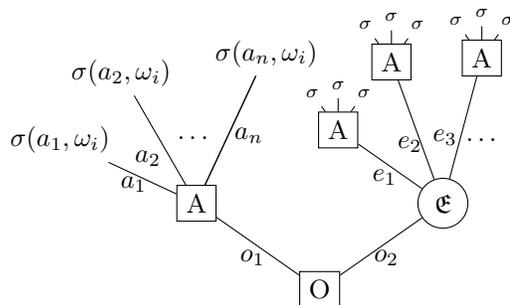


Figure 2: Illustration of Good’s Principle

At time t_0 you confront a decision problem $O = \{o_1, o_2\}$. If you implement option o_1 at time t_0 , then at time t_1 you will face a decision problem $A = \{a_1, \dots, a_n\}$ without observing the outcome of experiment \mathfrak{E} . If you implement option o_2 at time t_0 , then at time t_2 you will face the same decision problem

³Often a uniqueness result for probabilities and utilities accompanies the representation result (asserting, for example, that the probability function is unique and that the utility function is unique up to a positive affine transformation).

A after observing the outcome of experiment \mathfrak{E} . Now under the hypothesis N that at t_1 you face decision problem A after implementing option o_1 at t_0 (not observing the outcome of experiment \mathfrak{E}), let $\mathbf{c}(A|N)$ denote the set of admissible options given N (where \circ denotes functional composition):

$$\mathbf{c}(A|N) = \arg \max_{a \in A} \mathbb{E}_{p(\cdot|N)} \left[(u \circ \sigma)(a, p(d\omega|N)) \right].$$

Similarly, under the hypothesis K_i that at t_2 you face decision problem A after implementing option o_2 at t_0 and observing outcome e_i of experiment \mathfrak{E} at t_1 , let $\mathbf{c}(A|K_i)$ denote the set of admissible options given K_i :

$$\mathbf{c}(A|K_i) = \arg \max_{a \in A} \mathbb{E}_{p(\cdot|K_i)} \left[(u \circ \sigma)(a, p(d\omega|K_i)) \right].$$

Good’s principle assumes that at t_0 you are certain, regardless of whether or not you choose to observe the outcome of experiment \mathfrak{E} , that you will choose an option A which maximizes your expected utility, that your preferences over consequences remain unchanged, and that your beliefs given hypotheses accord with Bayesian conditionalization. Your expectation of (1) *your maximum conditional expected utility of choosing from A given experiment \mathfrak{E}* is not less than your expectation of (2) *your maximum conditional expected utility of choosing from A under option o_1* . That is, $o_2 \in \mathbf{c}(O)$, the set of admissible options from O. Moreover, your expectation of (1) is strictly greater than (2) unless there is an action from A that maximizes conditional expected utility from A regardless of the experimental outcome of \mathfrak{E} . In other words, unless the experiment is irrelevant, $\mathbf{c}(O) = \{o_2\}$.

4.2 Remarks on Conditional Probabilities

We wish to remark that your conditional probability judgments, whether precise or imprecise, concern only your commitments at the initial time t_0 . In our analysis we adopt a distinction made by Isaac Levi, and suggested, at least roughly, by many others.⁴ Specifically, we interpret your conditional probability judgments in one of two ways. First, according to the *called-off* interpretation, your conditional probability judgment given H expresses your commitment at time t_0 to specific *unconditional* attitudes contingent on the occurrence of H . According to de Finetti’s theory of previsions, for example, your conditional probability assessment of an event E given an event H at a particular time t_0 expresses your unconditional commitment at time t_0 to judge contracts concerning E that are “called-off” if H does not occur, where

⁴See, for example, [17, 18] and [23, 22, 10, 16, 34]; for a summary of Levi’s ideas, see [15, Appendix A].

they are posited to be nil. Alternatively, according to the *hypothetical* interpretation, your conditional probability judgment given H expresses your commitment at time t_0 to specific attitudes on the hypothetical supposition that H obtains. We contrast these two interpretations of conditional probability judgment with a third *temporal* interpretation which expresses your future commitment to attitudes upon observing that H obtains. In the sequential decision problems discussed in this paper, your current (at t_0) conditional probability judgments given a (possibly) future event H express your assessments on the hypothetical supposition that H is true. Similar remarks apply to other conditional judgments you endorse, such as your conditional value judgments and your conditional assessments of admissibility.

In our view, the question whether conglomerability is an appropriate normative standard for evaluating probability judgments in the senses of interest in this paper remains unsettled.

5 What’s so good about Good’s Principle?

Although Good’s Principle continues to be thought of as a cornerstone of orthodox Bayesianism by critics and champions alike, we maintain that the principle is not ironclad. In this section we consider two examples of violations of Good’s Principle in some detail, one involving dilation, another, non-conglomerability.

5.1 Good’s Principle and Dilation

Return to the dilation example from [28] that we began in §2.1. Recall that E is defined as the event of *either* the highly uncertain event G and the fair coin toss yielding the event H both occurring *or* both G and H failing to occur, that is $E := (G \cap H) \cup (G^c \cap H^c)$. Recall too that the probability of E and the probability of H are each determinate, namely $p(E) = \frac{1}{2} = p(H)$, whereas the probability of G is highly uncertain, namely $\overline{P}(G) = 0.9$ and $\underline{P}(G) = 0.1$.

Now suppose that at t_0 you face a decision problem $O = \{o_1, o_2\}$, where option o_1 is a **basic** decision problem A whereby you are to choose at t_1 between two acts: a_1 , which pays you \$1 if E occurs and ‘pays’ you $-\$1$ if E^c , i.e., $\sigma(a_1, E) = \$1$ and $\sigma(a_1, E^c) = -\$1$,⁵ or the act a_2 which ‘pays’ you a constant $-\$0.50$. Assume that your utility is linear in dollar amounts with $u(\$x) = x$. See Figure 3.

In this basic decision problem A, which is the result

⁵Here we abuse our notation by writing $\sigma(a, E) = \$1$, for instance, to express that $\sigma(a, \cdot)$ is a constant \$1 on E .

of implementing option o_1 , the subjective expected utility of a_1 is \$0 and the subjective expected utility of a_2 is $-\$0.50$. So, a_1 is uniquely admissible from A: receiving nothing is better than paying 50 cents.

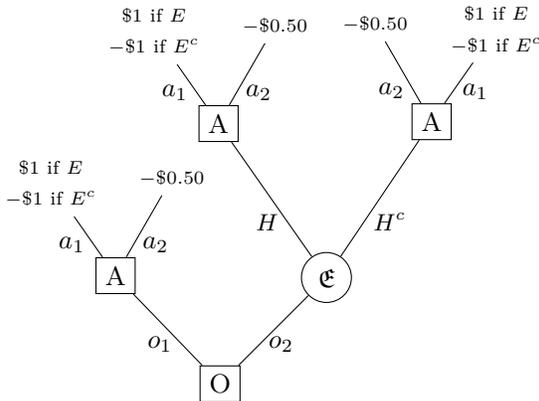


Figure 3: A Sequential Decision Example.

Turn now to option o_2 , whereby at t_2 you face a **derived** decision problem conditional on the outcome of experiment \mathfrak{E} . That is, you are confronted with the same decision problem A at t_2 after learning (only) that H obtains or H^c obtains at t_1 . But the derived decision problem A, which is the result of implementing option o_2 , is different from the basic decision problem A: in the derived decision problem the act a_1 is inadmissible against a_2 . Why? Because in the basic decision problem $p(E) = 1/2$, but in the derived decision problem E is dilated by \mathfrak{E} to 0.1 and 0.9: whether the outcome of the fair coin toss is heads or tails, E conditional on that outcome is highly uncertain. Thus, in the derived decision problem, there are probability mass functions $p \in \mathbb{P}$ whereby $p(E^c)$ is .9, in which case the minimum expected utility of a_1 is $-\$0.80$. So, in the derived decision problem, by Γ -Maximin, a_2 has a higher minimum expected value than a_1 regardless of the outcome of the experiment, \mathfrak{E} .

Assume that a decision maker is certain that she will not change her preferences, will update her belief state by Generalized Bayesian conditionalization, and that she will choose to maximize her minimal expected utility (Γ -Maximin). Then, in a pairwise choice between a_1 of the basic decision problem determined by option o_1 , which has an expected value of zero, and a_2 of the derived decision problem determined by option o_2 , which has an expected value of $-\$0.50$, observing cost-free information at t_1 , i.e., learning the outcome of the fair coin toss \mathfrak{E} , is devalued. Here we have a case where the decision maker would strictly prefer *not* to receive cost-free information!

Discussion. Although in finite spaces some decision rules, including Γ -Maximin, *require* decision makers to reject the opportunity to observe cost-free information before making a decision, others merely *permit* decision makers to reject the opportunity to observe cost-free information before making a decision. For example, E-Admissibility permits, but never requires, you to reject an opportunity to observe cost-free information before making a decision. Even so, E-admissibility supplemented with a secondary criterion for selecting among E-Admissible options—namely, to maximize expected utility with respect to a least informative distribution from among E-Admissible options—respects the value of (cost-free) information, and therefore mandates that decision makers abide by Good’s Principle. So, the first point to note is that dilating probabilities can be paired with a variety of decision rules, some abide by Good’s Principle, others do not.

The second point to emphasize is that E and H are *not* stochastically independent, so the basic (o_1) and derived (o_2) forms of the decision problem A are importantly different. (If the uncertainty in G were represented by a single probability rather than a set of probabilities, then the two forms would be equivalent.) From Theorem 1 we see that the association between E and H is the key to dilation; the effects one sees from evaluating conditional judgments merely are a consequence. Performing the experiment \mathfrak{E} reveals to you the extent of your uncertainty about the dependence of E on the experimental outcomes of \mathfrak{E} . How knowledge of this particular form of uncertainty affects decision making will depend on the decision maker’s beliefs, values and goals.

5.2 Good’s Principle and Non-Conglomerability

Suppose that at t_0 you face a decision problem $O = \{o_1, o_2\}$ as in the previous section, here with decision problem $A = \{a_1, a_2\}$ and experiment $\mathfrak{E} = \{H_n : n \in \mathbb{N}_{>0}\}$. Action a_1 pays you \$1 if E occurs and ‘pays’ you $-\$1$ if E^c , while action a_2 ‘pays’ you a constant $-\$0.50$ i.e., $\sigma(a_1, E) = \$1$ and $\sigma(a_1, E^c) = -\$1$ and $\sigma(a_2, E) = -\$0.50$ and $\sigma(a_2, E^c) = -\$0.50$.

Now, under the hypothesis that at t_1 you face the decision problem A without observing the outcome of experiment \mathfrak{E} , your subjective expected utility of a_1 is \$0, while your subjective expected utility of a_2 is $-\$0.50$. So you judge a_1 to be uniquely admissible from the **basic** decision problem A. That is, $\mathbf{c}(A|N) = \{a_1\}$, where N is the hypothesis that at t_1 you have implemented option o_1 .

Under the hypothesis that at t_2 you face the decision problem A after observing outcome H_i of \mathfrak{E} , your

subjective expected utility of a_1 is $-\$1$, while your subjective expected utility of a_2 remains $-\$0.50$. So you judge a_2 to be uniquely admissible from the **derived** decision problem A . That is, $\mathbf{c}(A|K_i) = \{a_2\}$, where K_i is the hypothesis that at t_2 you face the decision problem A after implementing option o_2 at t_0 and observing outcome e_i of experiment \mathfrak{E} at t_1 . Thus, assuming that you are certain you will not change your preferences, that you will update your belief state by Bayesian conditionalization, and that you will maximize subjective expected utility, option o_1 has constant utility $\$0$ and option o_2 has constant utility $-\$0.50$. In other words, $\mathbf{c}(A) = \{o_1\}$: at t_0 you judge that choosing from A without observing the outcome of \mathfrak{E} to be exclusively admissible for choice.

Discussion. One might argue that there are significant differences between failures of Good’s Principle due to dilation and failures due to non-conglomerability. For instance, in the case of dilation, some decision rules respect Good’s Principle and some do not, which has been cited as grounds for modifying particular decision rules rather than the modifying the uncertainty model. In the case of non-conglomerability, it may appear that there is a disanalogy. The standard reply to cases of non-conglomerability is to modify the uncertainty model, namely by imposing countable additivity, rather than to modify the expected utility maximization, which many take for granted. What are the grounds for adjudicating between these two cases?

6 Conclusion

In closing, there are three general points to make. First, notice that there are several familiar approaches that do not countenance imprecise probabilities but which nevertheless require decision makers to forgo the opportunity to observe cost-free information before making a decision, and some of those approaches do so even in finite spaces. Second, while Good’s Principle is often implicated in learning or sequential decision making, Good’s Principle itself is a synchronic, confirmational rule about an agent’s state of belief at a particular time, rather than a temporal rule regulating updating of an agent’s state of belief in light of an observation. Similarly, dilation is likewise characterized synchronically, rather than dynamically.

Finally, what is the normative standing of Good’s Principle? We believe it is not an obvious general principle of rationality, and that the classical argument strategies for establishing the principle rest on strong structural assumptions, not only about a decision maker’s adherence to expected utility maximization, but also about the decision maker’s beliefs about

her future preferences, future belief states, and future decision strategies. Although Good’s Principle is familiar, the foundations for its (still) wide acceptance are not; indeed, there appear to be a host of reasonable exceptions to Good’s Principle, even within the standard setting of utility maximization. For these reasons, we are puzzled why some authors still elevate Good’s Principle to a general normative principle while remaining indecisive about the normative status of (merely) finitely additive probabilities.

Acknowledgements

This research was supported in part by the Alexander von Humboldt Foundation.

References

- [1] Thomas E. Armstrong. Conglomerability of probability measures on boolean algebras. *Journal of Mathematical Analysis and Applications*, 150:335–358, 1990.
- [2] Inés Couso, Serafin Moral, and Peter Walley. Examples of independence for imprecise probabilities. In Gert de Cooman, editor, *Proceedings of the First Symposium on Imprecise Probabilities and Their Applications (ISIPTA)*, Ghent, Belgium, 1999.
- [3] Bruno de Finetti. Probability, statistics, and induction: Their relationship according to the various points of view. In Isotta Cesari and Leonard J. Savage, editors, *Probability, Induction, and Statistics, 1979*. John Wiley and Sons, 1959.
- [4] Bruno de Finetti. *Theory of Probability*, volume I. Wiley, 1990 edition, 1974.
- [5] Lester E. Dubins. Finitely additive conditional probability, conglomerability, and disintegrations. *Annals of Probability*, 3:89–99, 1975.
- [6] Adam Elga. Subjective probabilities should be sharp. *Philosophers Imprint*, 10(5), May 2010.
- [7] Gerd Gigerenzer and Henry Brighton. Homo heuristicus: Why biased minds make better inferences. *Topics in Cognitive Science*, 1:107–43, 2009.
- [8] I. J. Good. On the principle of total evidence. *The British Journal for the Philosophy of Science*, 17(4):319–321, 1967.
- [9] Peter Grünwald and Joseph Y. Halpern. When ignorance is bliss. In Joseph Y. Halpern, editor, *Proceedings of the 20th Conference on Uncertainty in Artificial Intelligence (UAI '04)*, pages 226–234, Arlington, Virginia, 2004. AUAI Press.

- [10] Ian Hacking. Slightly more realistic personal probability. *Philosophy of Science*, 34(4):311–325, 1967.
- [11] Timothy Herron, Teddy Seidenfeld, and Larry Wasserman. The extent of dilation of sets of probabilities and the asymptotics of robust bayesian inference. In *PSA 1994 Proceedings of the Biennial Meeting of the Philosophy of Science Association*, volume 1, pages 250–259, 1994.
- [12] Timothy Herron, Teddy Seidenfeld, and Larry Wasserman. Divisive conditioning: further results on dilation. *Philosophy of Science*, 64:411–444, 1997.
- [13] Brian Hill. Dynamic consistency and ambiguity: A reappraisal. Technical Report ECO/SCD-2013-983, HEC Paris, Paris, 2013.
- [14] Joseph B. Kadane, Mark J. Schervish, and Teddy Seidenfeld. Reasoning to a foregone conclusion. *Journal of the American Statistical Association*, 91(435):1228–1235, 1996.
- [15] Joseph B. Kadane, Mark J. Schervish, and Teddy Seidenfeld. Is ignorance bliss? *Journal of Philosophy*, 105(1):5–36, 2008.
- [16] Henry E. Kyburg, Jr. Bets and beliefs. *American Philosophical Quarterly*, 1:54–63, 1968.
- [17] Isaac Levi. Confirmational conditionalization. *Journal of Philosophy*, 75(12):730–737, 1978.
- [18] Isaac Levi. *The Enterprise of Knowledge*. MIT Press, Cambridge, MA, 1980.
- [19] David V. Lindley. *Introduction to Probability and Statistics*. Cambridge University Press, Cambridge, 1965.
- [20] Mark J. Machina. Dynamic consistency and non-expected utility models of choice under uncertainty. *Journal of Economic Literature*, 27(4):1622–68, 1989.
- [21] Arthur Paul Pedersen and Gregory Wheeler. Demystifying dilation. *Erkenntnis*, 79(6):1305–1342, 2014.
- [22] Howard Raiffa and Robert Schlaiffer. *Applied Statistical Decision Theory*, volume 1 of *Studies in Managerial Economics*. Harvard Business School Publications, 1961.
- [23] Frank P. Ramsey. Truth and probability. In H. E. Kyburg and H. E. Smokler, editors, *Studies in subjective probability*, pages 23–52. Robert E. Krieger Publishing Company, Huntington, New York, second (1980) edition, 1926.
- [24] Frank P. Ramsey. *The Foundations of Mathematics and Other Essays*, volume 1. Humanities Press, New York, 1931.
- [25] Leonard J. Savage. *Foundations of Statistics*. Dover, New York, 1972.
- [26] Mark J. Schervish, Teddy Seidenfeld, and Joseph B. Kadane. The extent of non-conglomerability of finitely additive probabilities. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, 66(2):205–226, 1984.
- [27] Mark J. Schervish, Teddy Seidenfeld, and Joseph B. Kadane. On the equivalence of conglomerability and disintegrability for unbounded random variables. *Statistical Methods and Applications*, 23:501–518, 2014.
- [28] Teddy Seidenfeld. When normal and extensive form decisions differ. In D. Prawitz, B. Skyrms, and D. Westerstahl, editors, *Logic, Methodology and Philosophy of Science*. Elsevier Science B. V., 1994.
- [29] Teddy Seidenfeld. A contrast between two decision rules for use with (convex) sets of probabilities: Γ -Maxmin versus E-admissibility. *Synthese*, 140:69–88, 2004.
- [30] Teddy Seidenfeld, Mark J. Schervish, and Joseph B. Kadane. Non-conglomerability for finite-valued, finitely additive probability. *he Indian Journal of Statistics, Series A*, 60(3):476 – 491, 1998.
- [31] Teddy Seidenfeld and Larry Wasserman. Dilation for sets of probabilities. *The Annals of Statistics*, 21:1139–154, 1993.
- [32] Marciano Siniscalchi. Dynamic choice under ambiguity. *Theoretical Economics*, 6:379–421, 2011.
- [33] Peter Wakker. Nonexpected utility as aversion of information. *Journal of Behavioral Decision Making*, 1:169–75, 1988.
- [34] Peter Walley. *Statistical Reasoning with Imprecise Probabilities*. Chapman and Hall, London, 1991.
- [35] Larry Wasserman and Teddy Seidenfeld. The dilation phenomenon in robust Bayesian inference. *Journal of Statistical Planning and Inference*, 40:345–356, 1994.
- [36] Marco Zaffalon and Enrique Miranda. Probability and time. *Artificial Intelligence*, 198:1–51, 2013.