心理科学进展 2011, Vol. 19, No. 5, 617–623
Advances in Psychological Science DOI: 10.3724/SP.J.1042.2011.00617

• 主编特邀(Editor-In-Chief Invited) •

编者按：

本期“主编特邀”的作者是法国心理学家 Jean-François Bonnefon, 现任法国国家科学研究中心(CNRS)研究员, 法国图卢兹大学 CLLE (Cognition, Langues, Langage et Ergonomie)研究所所长。Bonnefon 的研究领域涉及两个学科—— 心理学/语言学, 以及计算机科学/人工智能。他主要针对理性心理活动的三剑客—— 推理、判断和决策——进行了大量的研究工作, 并在 Psychological Science、Psychological Review 等学术期刊上发表过论文 40 余篇。本文中, Bonnefon 将介绍一个在社会科学领域已引起广泛关注, 但却缺乏行为心理学研究的判断与决策问题—— 教条悖论。该悖论普遍存在于群体决策和判断整合过程之中。他通过总结以往的研究结果、分析初步的实证数据, 对这一问题的研究意义、研究现状和发展方向进行了深入探讨。相信本文将给关心判断与决策领域的心理学者带来新的研究思路。

(本文责任编辑：李纾)

The Doctrinal Paradox, a New Challenge for Behavioral Psychologists

Jean-François Bonnefon

(*Centre National de la Recherche Scientifique Université de Toulouse, France*)

Abstract: In various professional and private contexts, it is often necessary to aggregate different opinions about whether a given claim is true or false. A doctrinal paradox occurs when this claim is akin to a logical formula combining several propositions, and it turns out that the claim itself is true (resp., false) for a majority of judges, whereas a majority of judges has opinions on the propositions that would make the claim false (resp., true). The doctrinal paradox is a serious formal concern for judgment aggregation, which has generated intense normative research in various scientific fields. Behavioral psychologists, though, still have to undertake systematic research on this important problem. This article provides a brief introduction to the doctrinal paradox and its formal study, summarizes available behavioral data, and points to perspective for future behavioral research.

Key words: opinion; aggregation; paradox; vote

1 Introduction

Imagine that you are considering buying a car, which the salesperson described as being safe and having low maintenance costs. Because you are not sure that this description is correct, you ask five knowledgeable persons to tell you whether or not this specific car is safe, and whether or not it has low maintenance costs. You then summarize their responses in a table, which looks just as Table 1. The rule you gave to yourself was that you would consider the description as correct if and only if a majority of experts had opinions that agreed with that description. From the set of opinions displayed in Table 1, what do you conclude about the description?

Quite remarkably, you might conclude anything you want. You might try to count the number of experts who agree with the description as

Received date: 2011-01-26
Address for correspondence: Jean-François Bonnefon (Cognition, Langues, Langage et Ergonomie) Maison de la recherche, 5 allées A. Machado, 31058 Toulouse Cedex 9, France. Telephone: 33 5 61 50 36 01. Email: bonnefon@ univ-tlse2.fr

a whole, that is, who see the car as safe and having low maintenance costs. There are only two of them, out of five experts. As a consequence, you would conclude that the description is incorrect. But consider now this other perspective: three experts (out of five) agree that the car is safe, so that is a majority; and three experts (out of five) agree that the car has low maintenance costs, and again that is a majority. As a consequence, you have legitimate grounds to conclude that the description was correct.

Table 1: A set of opinions leading to a doctrinal paradox: a majority of experts say it is true that the car is safe, a majority say it is true that the car has low maintenance cost, but only a minority of experts say that both things are true.

	Safe car?	Low costs?	*Both?*
Expert 1	Yes	No	*No*
Expert 2	No	Yes	*No*
Expert 3	Yes	Yes	*Yes*
Expert 4	Yes	Yes	*Yes*
Expert 5	No	No	*No*

This is a remarkable situation: depending on the way you count, the same set of opinions seems to support equally well two opposite conclusions. There is a clear majority in favor of one conclusion, but also a clear majority in favor of the opposite conclusion. What should you conclude in such a situation? And, perhaps more importantly from the point of view of the psychologist, what are you *actually* going to conclude?

As we will see in Section 2, this example illustrates a situation known as the doctrinal paradox, which is itself a key problem in the field of judgment aggregation. Judgment aggregation and its paradoxes (doctrinal and otherwise) are currently generating much excitement in the social and computer sciences, as we will see in Section 3. The behavioral sciences, however, seem to be late in the game, and there is very little data yet on the way laypersons escape the doctrinal paradox. We will survey these data in Section 4, and conclude on the need for psychologists to tackle this new challenge.

2 The doctrinal paradox in judgment aggregation

Doctrinal paradoxes can occur when aggregating the opinions of a group of individuals about a set of logically connected propositions. In these situations, the individuals within the group all expressed a set of binary opinions (true/false) about a series of propositions *x*, *y*, *z*, etc. Our car example features two propositions *x* and *y*, which respectively stand for “the car is safe” and “the car has low maintenance costs.”

The goal of the aggregation is to define the collective opinion of the group about a logical formula combining these propositions. In our example, the formula to be assessed is “*x* and *y*”, that is “the car is safe and it has low maintenance costs.” In other situations, we could be interested in other simple formulas, such as the disjunction “*x* or *y*”, or in more complex formulas featuring more variables, such as “*x* or not-(*y* and *z*).”

There are many situations in which we might want to aggregate opinions this way. For example, current attempts at harnessing the wisdom of the crowds through social networks have to address the challenge of opinion aggregation. List and Polak (2010), in their introduction to a recent special section of the *Journal of Economic Theory* devoted to judgment aggregation, describe a number of other relevant contexts, such as courts of laws, scientific panels, or search committees. For example, an academic department could be looking for candidates with outstanding teaching evaluations (*x*), an excellent track record of publications (*y*), and some experience in attracting research funding (*z*). What is expected from the search committee is an aggregated opinion on whether each candidate matches this description. The description itself is a logical formula combining the three propositions, namely “*x* and *y* and *z*.”

There are two rather natural ways to calculate the collective opinion of the group about such a formula, based on individual sets of judgments about

its propositions. First, one might simply count how many individuals expressed sets of judgments which make the formula true. In the search committee example, this would amount to counting the number of committee members who believed that *x*, *y*, and *z* were all true about a given candidate. If they are the majority, then the committee is seen as collectively agreeing with this description of the candidate. This procedure is often called *conclusion-based.*

The other option is to count how many individuals agree with each proposition in the formula. Once this is done, one can check whether the formula is true when each proposition is given the value so aggregated. In the search committee example, this would amount to counting how many members agree with *x*, then how many members agree with *y*, then how many members agree with *z*. Each proposition is given the aggregated value True if a majority of members agreed with it, and the aggregated value False otherwise. The committee is seen as collectively agreeing with the formula if the aggregated values of the propositions make the formula true. This procedure is often called *premise-based.*

A doctrinal paradox occurs when these two procedures give inconsistent results. Let us check that this is the case in our car example. The conclusion-based procedure amounts to counting how many experts answered Yes to both questions. Only two of them did, so this procedure tells us that the five experts collectively see the description as false. The premise-based procedure, on the contrary, tells us that the aggregated value of the first proposition is True (three yesses), and that the aggregated value of the second proposition is also True (again, three yesses). These aggregated values make the formula true, so this procedure tells us that the experts collectively see the description as true. The two procedures thus deliver contradictory results, leaving us in the midst of a doctrinal paradox.

3 Formal results

The doctrinal paradox, as well as other related problems, provided the impetus for the vibrant field of judgment aggregation, generating research in law, political science, economics, philosophy, and computer science (Bovens & Rabinowicz, 2006, Cariani, Pauly, & Snyder, 2008, Dietrich, 2006, Dietrich & List, 2008, List, 2005, List & Pettit, 2002, 2004, Pigozzi, 2006).

A substantial part of this research is inspired by the similarities between judgment aggregation and the more classic problem of preference aggregation. Because the present article is meant as an entry point for psychologists, we will not consider here the important but rather abstract results that were obtained by drawing on these similarities.[1] We will, however, briefly consider two questions that are of substantial interest to psychologists, and that were formally explored by analytical scientists: How likely is it that a doctrinal

1 Judgment aggregation is a more general situation than preference aggregation, and the doctrinal paradox itself is a generalization of Condorcet's paradox. Because judgment aggregation is a generalization of preference aggregation, the constraints on the input and output of the aggregation procedure can go beyond the constraints that are typically considered in preference aggregation. For example, instead of asking for transitivity of the individual and collective preference orderings, judgment aggregation can require that both the individual and collective sets of judgments are logically consistent. The formal literature on judgment aggregation has produced impossibility results akin to Arrow's theorem, and has offered results concerning the requirements that can be lifted in various situations in order to obtain a compelling aggregation procedure. Because judgment aggregation can operate on non-evaluative judgments instead of preferences, it makes it especially relevant to investigate the truth-tracking capacity of various aggregation procedure, that is, to assess the probability that a given procedure will yield an objectively correct judgment in a given situation. For an entry point in that literature, see List and Polak (2010).

paradox will appear in real life? And what procedure gives the best results when it does?

List (2005) developed a model for determining the probability of a doctrinal paradox, under some assumptions about the distribution of individual sets of judgments. The model considers a group of *n* individuals, who all give their opinion about the truth or falsity of the two variables *x* and *y*. Each individual can thus communicate one of four sets of judgments (True-True, True-False, False-True, False-False). Each of these sets of judgments has a fixed probability of being communicated, which is the same for all individuals. Finally, individual sets of judgments are assumed to be independent from each other.

When the four sets of judgments have an equal probability, the probability of a doctrinal paradox quickly converges to .25 as *n* increases. The probability of a doctrinal paradox can even be much higher under some conditions. For example, it quickly converges to 1 in a situation when False-False judgments are highly unlikely (p=.01) and all other judgments are equally likely (p=.33). For small groups (say, from n=11 to n=31) paradoxes occur with a probability ranging from .20 to .90 in the various scenarios considered to illustrate the model.

It would thus appear that doctrinal paradoxes can be obtained very easily when aggregating the opinions of a group. In other words, premise-based aggregation and conclusion-based aggregation can easily deliver inconsistent responses. The next question, then, is to consider which of these procedures delivers the *better* answer. This can be done by simulating situations where the correct answer is known, and to check which procedure has the highest probability to deliver this correct answer. While this question might not be entirely solved, an array of results would suggest that the premise-based procedure should often be preferred (e.g., Hartmann, Pigozzi, & Sprenger, 2010). That is not to say, however, that laypersons spontaneously turn to the premise-based procedure when they face a doctrinal paradox. This descriptive issue is addressed in the next section.

4 Behavioral data

There is very little data available on how people behave when facing a doctrinal paradox, or more generally about whether they prefer premise or conclusion based procedure for aggregating opinions. Interestingly, some data about jury behavior seem to suggest that jurors benefit from using a premise-based procedure. The analogy here is that the facts about the case are the premises, and the verdict is the conclusion. Conclusion-based aggregation amounts to aggregating opinions about the verdict, while premise-based aggregation amounts to aggregating opinions about the facts, and only then to calculate the verdict. Juries that primarily try to aggregate opinions about the facts, rather than about the verdict, seem to do a more careful work and to be more satisfied with their experience (Hastie, Penrod, & Pennington, 1983). Furthermore, mocked juries who are encouraged to deliberate in a premise-based way (by aggregating their opinions before members have individually committed to a verdict) reach decisions that all members find more satisfying, even members who actually disagree with the final decision (Kameda, 1991).

It would thus appear that the premise-based procedure is both a formally and psychologically adequate escape route from the doctrinal paradox. The problem, though, is that people do not seem to manifest a spontaneous preference for that solution, as shown by the two experimental papers that directly addressed individual behavior in a situation of doctrinal paradox (Bonnefon, 2007, 2010).

Bonnefon (2007) reported that people manifested a slight preference for the *conclusion-based* procedure, apparently for reasons of simplicity: They found the conclusion-based procedure simpler, and these judgments of simplicity mediated their preference. This

preference, however, could be attenuated by various experimental manipulations. The premise-based procedure became more appealing when the variables *x* and *y* were negatively correlated in the real world. It seems that when subject did not expect *x* and *y* to be true together, they satisficed with the fact that *x* was true for a majority and *y* was true for a majority, without requiring that a majority actually believed that *x* and *y* were true at the same time. The cover story in the experiment was one where employees were judged on two criteria. In both the experimental conditions, *x* stayed the same: "being young". The criterion *y*, however, was manipulated to be either independent from *x* ("being trilingual"), or negatively correlated with *x* ("having extensive high-level managerial experience"). Faced with structurally identical sets of judgments, participants preferred conclusion-based aggregation when looking for someone young and trilingual, but they preferred premise-based aggregation when looking for someone young with extensive, high-level managerial experience.

The premise-based procedure can also be promoted by specific framings of the aggregation task. Let us look again at the scenario when the goal of the aggregation is to find an employee who is young (*x*) and trilingual (*y*). The set of available judgments exhibits a doctrinal paradox, as the premise-based procedure says the employee is young and trilingual, whereas the conclusion-based procedure says he is not. Now let us add one element to the situation; an employee who is found to be young and trilingual will be moved to either a very coveted position, or to a position that nobody wants to fill.

When this element is added to the scenario, it changes the preferences of subjects about how judgments should be aggregated. Subjects are more likely to find the premise-based procedure appropriate when its output would be positive for the employee (Bonnefon, 2007). This result was explored further in another pair of experiments (Bonnefon, 2010), in which the framing of the aggregation was manipulated slightly differently. In these experiments, the goal of the aggregation was to assess whether employees where "competent" and "motivated" (in one condition) or "incompetent" and "unmotivated" (in another condition). Importantly, this variable was manipulated within-subject. That is, the same subjects saw two identical sets of judgments, only with different labels for the criteria.

As it turned out, the labels for the criteria made a large difference. When the criteria were labeled "competent" and "motivated", subjects preferred the premise-based procedure, which said the employee was both of these things. When the criteria were labeled "incompetent" and "unmotivated", the same subjects, given structurally identical sets of judgments, now preferred the conclusion-based procedure, which said that the employee was not both of these things.

A second experiment used a disjunctive variant of the paradox, where the goal of the aggregation was to assess whether employees where "competent" *or* "motivated" (in one condition) or "incompetent" *or* "unmotivated" (in another condition). Once more, this variable was manipulated within-subject. The important aspect of this experiment is that the conclusion-based procedure now says that the employee is (in)competent or (un)motivated, whereas the premise-based procedure says that he is not. This time, subjects preferred the conclusion-based procedure when the labels were "competent" and "motivated", but the premise-based procedure when the labels were "incompetent" and "unmotivated".

This overall pattern of results appears to suggest that the framing of the paradox affects the type of mistakes people feel concerned about, and that people then adopt the procedure that is the most likely to allay the specific concern they feel. When the goal of the aggregation is to decide whether someone is competent and/or motivated, excluding someone by mistake seems to be more of a concern than including someone by mistake. Conversely,

when the goal of the aggregation is to decide whether people are incompetent and/or unmotivated, including someone by mistake seems to be more of a concern than excluding someone by mistake.

As it turns out, for the classic (conjunctive) version of the paradox, the premise-based procedure minimizes the risk of excluding someone by mistake, whereas the conclusion-based procedure minimizes the risk of including someone by mistake. For the disjunctive version of the paradox, the two aggregation rules have opposite properties (List, 2006). Thus, all happens as if subjects were using the procedure that is the most effective at allaying the most salient concern raised by the framing of the aggregation. Rather than having a fixed preference for premise-based or conclusion-based aggregation, people might pick that procedure which is the most likely to avoid errors of omission or errors of commission, as a function of the error they wish to avoid the most. This is an encouraging perspective, as it would suggest that people are not totally confused by situations of doctrinal paradox, but may rather have some intuitive grasp of the properties of the aggregation procedure they end up using.

5 Perspectives for future behavioral research

The doctrinal paradox raises a deep problem for judgment aggregation, as soon as we wish to reach a collective judgment about a complex topic. When we wish to reach a collective judgment about an issue that can be defined as a logical formula combining several propositions, we can frequently be in a situation where aggregating the judgments about the propositions delivers a different result than aggregating the judgments about the formula itself. Note that this situation is quite likely to arise in situations when people take advice from multiple advisors. Although multiple-advisor situations have been addressed in the advice giving literature (see Bonaccio & Dalal, 2006, p. 137, for a review), advice-taking research has not yet linked with the doctrinal paradox literature.

This problem has generated a large body of normative research, addressing the issue of what should be done when a doctrinal paradox arises. There is comparatively much less research, however, on what people actually do in such a situation. This is a gap that behavioral psychologists must fill. Behavioral research is needed in particular to ease the transition from normative findings to prescriptive recommendations: Knowing what should be done is only one step to telling people what to do, as it is also necessary to know what they would be spontaneously inclined to do and why.

Behavioral research is also needed to protect people from manipulation. A worrying aspect of the doctrinal paradox is that whomever can decide on the procedure can decide on the outcome of the aggregation. If psychologists want to prevent that outcome, they must assess the extent to which people understand situations of doctrinal paradox,[2] and the extent to which they can control their reaction to such factors as the way the situation is framed.

In the previous section, we have considered a rather optimistic perspective on framing effects as applied to the doctrinal paradox. These framing effects might reflect a rational goal of minimizing the type of errors that people mostly want to avoid in different situations. It is not clear, however, whether this rational strategy would be deliberate, or automatically triggered by surface properties of the situation. In that regard, it is perhaps worrying that Bonnefon (2010) did not find any moderation of the framing effect by individual characteristics such as the Need for Cognition. It remains to be

2 The extent to which people can spontaneously understand doctrinal paradoxes is still largely unknown. In the context of piloting experimental materials, this author realized that many subjects appeared to be unable to properly apply premise-based aggregation when required to.

seen whether such effects or others can be moderated by other individual characteristics, or influenced by cultural differences. Lastly, behavioral research on the doctrinal paradox will have to go beyond the framing effects demonstrated by Bonnefon (2007, 2010), and identify others psychologically relevant aspects of judgment aggregation that may swing people's choice of an aggregation procedure when they try to come to terms with the inconsistent conclusions they can reach from a single set of judgments.

Refferences

Bonaccio, S., & Dalal, R. (2006). Advice taking and decision-making: An integrative literature review, and implications for the organizational sciences. *Organizational Behavior and Human Decision Processes, 101*, 127–151.

Bonnefon, J. F. (2007). How do individuals solve the doctrinal paradox in collective decisions? An empirical study. *Psychological Science, 18*, 753–755.

Bonnefon, J. F. (2010). Behavioral evidence for framing effects in the resolution of the doctrinal paradox. *Social Choice and Welfare, 34*, 631–641.

Bovens, L., & Rabinowicz, W. (2006). Democratic answers to complex questions – an epistemic perspective. *Synthese, 150*, 131–153.

Cariani, F., Pauly, M., & Snyder, J. (2008). Decision framing in judgment aggregation. *Synthese, 163*, 1–24.

Dietrich, F. (2006). Judgment aggregation: (im)possibility theorems. *Journal of Economic Theory, 126*, 286–298.

Dietrich, F., & List, C. (2008). Judgment aggregation without full rationality. *Social Choice and Welfare, 31*, 15–39.

Hartmann, S., Pigozzi, G., & Sprenger, J. (2010). Judgment aggregation and the problem of tracking the truth. *Journal of Logic and Computation, 20*, 603–617.

Hastie, R., Penrod, S., & Pennington, N. (1983). *Inside the jury*. Cambridge, MA: Cambridge University Press.

Kameda, T. (1991). Procedural influence in small-group decision making: Deliberation style and assigned decision rule. *Journal of Personality and Social Psychology, 61*, 245–256.

List, C. (2005). The probability of inconsistencies in complex collective decisions. *Social Choice and Welfare, 24*, 3–32.

List, C. (2006). The discursive dilemma and public reason. *Ethics, 116*, 362–402.

List, C., & Pettit, P. (2002). Aggregating judgments: An impossibility result. *Economics and Philosophy, 18*, 89–110.

List, C., & Pettit, P. (2004). Aggregating judgments: Two impossibility results compared. *Synthese, 140*, 207–235.

List, C., & Polak, B. (2010). Introduction to judgment aggregation. *Journal of Economic Theory, 145*, 441–466.

Pigozzi, G. (2006). Belief merging and the discursive dilemma: an argument-based account to paradoxes of judgment aggregation. *Synthese, 152*, 285–298.

教条悖论：行为心理学家的新挑战

Jean-François Bonnefon

(法国国家科学研究中心; 图卢兹大学, 法国)

摘 要 在各种专业或私人情境中, 人们经常需要整合不同的意见来判断某一观点的对错。当观点的形式类似于将多个论点组合而成的逻辑公式(使用"且"、"或"等逻辑连接词)时, 容易产生教条悖论(doctrinal paradox)。即, 虽然整体意见支持该观点是正确的(或错误的), 但是分析这些整体意见中所包含的各个论点, 却会得到相反的结论。教条悖论是判断整合研究中关注的重要问题, 已在各科学领域引发大量的规范性研究。行为心理学家需在这个重要问题上开展系统研究。本文简要介绍了教条悖论及其过往研究, 总结已有的行为数据, 并指出未来行为研究的方向和视角。

关键词 观点; 整合; 悖论; 表决

分类号 B842

(中文题目、摘要翻译: 杜雪蕾, 饶俪琳)