Theory choice with ordinal grading

(or: Kuhn meets Majority Judgment)

Thomas Boyer-Kassem *

March 2026

Abstract

How should scientists choose between rival theories? Kuhn observed that theory choice relies on multiple criteria, whose imprecision and variable weighting may lead to different yet rational choices. Recent work has modeled this problem using social choice theory, yielding Arrow-type impossibility results. I propose a different framework in which scientists assign qualitative ordinal grades, rather than rankings, to theories on each criterion. This richer informational basis enables an axiomatic approach with an aggregation using Majority Judgment at individual and collective levels. The framework thus provides a formal account of Kuhn's "no unique algorithm" thesis.

1 Introduction

How should scientists choose between several rival theories? This question has been famously studied by Kuhn (1977), who observed that scientists rely on several criteria of theory choice, or "values", such as accuracy, simplicity, consistency, scope, and fruitfulness. Difficulties arise from the use of such values: they are imprecise — i.e. their interpretations can vary across scientists — and they may conflict with one another, leading scientists to assign them different weights. Kuhn therefore argued that different theory choices can legitimately be regarded as equally rational — his "no unique algorithm" thesis. More recently, a debate has emerged over whether theory choice can be formally modeled as a social choice problem, leading to Arrow's impossibility result (see, for instance, Morreau 2015; Okasha 2011, 2015). In this analogy, criteria play the role

*Université de Poitiers, MAPP, Département de Philosophie, 36 rue de la Chaîne, F-86000 Poitiers, France. E-mail: thomas.boyer.kassem@univ-poitiers.fr

of voters, and the input consists of ranking of theories according to the various criteria. However, rankings are not very informative: a theory may rank first on a criterion by satisfying it very strongly or only slightly, and the second and third theories may be either very close or very far apart on the underlying scale. It has been noted that Arrow's impossibility can be avoided by moving beyond rankings and enriching the informational basis (Okasha 2011). This paper pursues this idea and proposes a different modeling framework: scientists assign qualitative judgments to each theory on each criterion using a common ordinal scale. For instance, instead of saying that, with respect to simplicity, Copernicus' theory ranks first and Ptolemy's second, one might say that the Copernicus' theory is *very good* in simplicity and Ptolemy's only *fair*. Such qualitative evaluations are both natural and familiar: scientists can readily provide them when assessing theories, and similar grading practices are standard in grant evaluation and peer review.

This raises the following question: How should scientists choose the best scientific theory when several criteria are assessed using an ordinal scale? This question has two dimensions. The first is individual, when a given scientist chooses which theory to accept or pursue. This issue is addressed in Section 2. The second is collective, when the evaluations of a scientific community (for instance as reconstructed by a historian) should be aggregated. This issue is addressed in Section 3. Using an axiomatic approach, I argue that in many cases at the individual level, the suitable rule, i.e. a rule which satisfies certain epistemic desiderata, is Majority Judgment, an algorithm developed in social choice theory to rank wines and political candidates. Under this richer information framework, Arrow's impossibility gives way to a possibility result. At the collective level, scientists' evaluations should again be aggregated using Majority Judgment, though two distinct procedures are available, each with its own advantages. Overall, the paper proposes a novel modeling framework for theory choice that helps clarify Kuhn's analysis and yields normative results at both the individual and collective levels.

2 Theory choice at the individual level

Even though scientists belong to a community, the decision to accept or pursue a particular scientific theory is ultimately an individual one. In this section, I examine how this decision should be made, given the qualitative judgments a scientist assigns to rival theories on several criteria and the weights assigned to these criteria.

2.1 Framework

The framework I consider is the following. A scientist faces m rival scientific theories, and considers p criteria of theory choice, or epistemic values (Kuhn's typical list, recalled in the Introduction, has $p = 5$). These criteria may have unequal weights, which sum to 1. I introduce a novel, central hypothesis:

> **(Hypothesis) Ordinal Grading:** There exists a finite set of grades (e.g. *Excellent*, *Very Good*, *Good*, *Fair*, *Poor*) arranged as an ordinal scale with a total order. For each theory and each criterion, the scientist assigns a grade from this same scale.

An ordinal scale with a total order means that the grades can be ordered from lowest to highest (they can be meaningfully compared by "equal to", "greater than" or "less than", cf. Balinski and Laraki 2010 chap. 8; Morreau 2016), but it is meaningless to add, subtract or average the grades. For instance, which grades might an astronomer have assigned to the theories of Ptolemy and Copernicus at the turn of the 17th-century? According to Kuhn (1977, p. 323), they could not be distinguished in terms of accuracy, suggesting the same grade, *Good*. Copernicus' theory conflicted with the received physical theory, and might therefore receive a *Bad* for consistency, while Ptomely's would receive *Good*. As for simplicity, it "favored Copernicus" (Kuhn 1977, p. 324), if evaluated by the amount of mathematical apparatus needed to explain the gross qualitative features of celestial bodies. The astronomer might thus assign *Very Good* to Copernicus and *Fair* to Ptolemy for simplicity.

Assigning qualitative judgments in terms of grades is not a difficult task for a scientist, who is assumed to have some knowledge of the competing theories. One might nevertheless object that, if rankings are abandoned, ordinal grading may not be the best substitute — why not use an interval or a ratio scale? This direction is partly explored by Okasha (2011), who argues that Bayesianism provides a framework in which accuracy can be measured on a ratio scale. However, he also acknowledges that not all criteria can be measured this way; for instance, fruitfulness is to be measured by an ordinal scale. Overall, my framework assumes only that ordinal grading is meaningful; it is not problematic if some criteria allow finer assessments than these grades. In the special case where scientists consider only criteria measurable on a ratio scale, the probability aggregation literature (see Dietrich and List 2016 for a review) should be used instead of the present approach. How, then, should our astronomer rank the theories of Ptolemy and Copernicus based on the grades assigned?

2.2 An axiomatic approach

To determine which ranking rule[1] is best, it is standard to adopt an axiomatic approach: theoretical desiderata (axioms) are formulated, and one then examines which ranking rules, if any, satisfy them. Here, the axioms are epistemic. I adapt them from the political axioms introduced in social choice theory by Balinski and Laraki (2020). I write $T_A \succeq T_B$ when the ranking rule places theory T_A above theory T_B. The first axiom directly follows from the Ordinal Grading hypothesis:

> **(Axiom 1) Grades.** The ranking rule takes as input the $m \times p$ grades assgined by the scientist to the theories on all criteria.

Implicit in this axiom is that the ranking rule does not take rankings of theories as input, which is the main difference with Arrow's (1951) or Okasha's (2011) approaches.

Axiom 2 concerns the input domain of the ranking rule. It states that the grades a scientist can assign are not constrained *a priori*:

> **(Axiom 2) Unrestricted Domain.** Grades may be assigned from the scale without restriction.

That is, the ranking rule should be able to accept any set of grades as input. One might object, in the spirit of Morreau's (2015) critique of Okasha (2011), that some criteria of theory choice are rigid, i.e., it is a metaphysical impossibility for theory T_A to be simpler, say, than T_B, which would make requiring Axiom 2 unsound. However, Axiom 2 can be given a pragmatic rather than metaphysical reading. The requirement is that the same ranking rule be usable across many situations, for instance irrespective of the particular scientist assigning grades, the theories considered, their number, or the criteria employed. Even if grades were fixed in some metaphysical sense in a given case, they are not fixed across an unrestricted range of situations. Otherwise, it would be odd to tell scientists: "Because you're not considering simplicity, you should switch to another ranking rule." It is simpler and more robust to have one-size-fits-all aggregation rules. This differs from Arrow's (1951) more general framework (cf. Morreau 2015, p. 242): in theory choice, it may be reasonable *not* to require that the set of rival theories and the set of criteria remain fixed. Moreover, even if the rigidity objection succeeded and Axiom 2 had to be dropped, the consequences would be less severe than in Okasha's framework. In his

[1] Technically, a ranking rule orders all rival theories from best to worst. One may also consider an aggregation rule, which selects the best theory, or a grading rule, which assigns a grade to each theory. What follows about ranking rules can be adapted to aggregation rules or grading rules; in particular, Majority Judgment in Section 2.3 yields all three.

case, the axiom is required to derive an analogue of Arrow's impossibility result. Here, it serves as a sufficient condition for a possibility result (see below); dropping it would merely remove the uniqueness characterization — Majority Judgment (introduced below) would remain a suitable aggregation rule, though no longer proven to be the only one.

Next, Axiom 3 is standard:

> **(Axiom 3) Anonymity.** Permuting the criteria (together with their weights) does not change the outcome.

The permutation should be understood formally, as referring to the indices assigned to criteria in the mathematical function. The axiom states that the ranking rule should not yield a different result just because, say, simplicity appears as the second argument of the function rather than the first.[2] A similar permutation requirement concerns the theories:

> **(Axiom 4) Neutrality.** Permuting the theories does not change the outcome.

Next, a requirement concerns how grades are taken into account: better epistemic assessments should matter, which is hardly controversial:

> **(Axiom 5) Monotonicity.** If $T_A \succeq T_B$ and the grade given to T_A for one (or more) criterion is increased, then $T_A \succ T_B$.

(where $T_A \succ T_B$ means $T_A \succeq T_B$ and not $T_B \succeq T_A$). Axiom 6 states that the ranking rule should always deliver a ranking:

> **(Axiom 6) Completeness.** For any pair of theories (T_A, T_B), either $T_A \succeq T_B$ or $T_B \succeq T_A$.

This desideratum seems natural: we are considering a situation in which the scientist is supposed to choose between rival theories by ranking them. One objection has to be considered, though, since the desirability of this axiom depends on the nature of the choice. If the choice concerns pursuing a theory, it is reasonable to assume that the scientist must decide between them (in the Kuhnian picture for instance, no scientist simultaneously works within two competing paradigms). In that case, the ranking rule should indeed rank the theories, as Axiom 6 requires. By contrast, if the choice concerns teaching theories to students, or funding research projects associated with rival theories, it may be

[2] One possibility Kuhn did not consider is that some criteria might be lexicographically ordered — for instance, theories could be ranked first by accuracy, with other weighted criteria invoked only to break ties. In such a case, the present framework would apply solely to this second stage, where the remaining criteria are compared by weight.

possible to refuse to choose: one might teach two competing theories during a semester, or divide available funding. In Kuhn's analysis, such a refusal to rank theories need not arise from equality but from incommensurability — for instance, if the theories excel on different criteria and selecting one would entail a loss that cannot be compensated. One might then simply refuse to rank them and insist that both deserve to be taught or funded. Overall, Axiom 6 is a natural requirement for individual decisions about the pursuitworthiness of theories (Kuhn's original problem), and may be questioned only for decisions about teaching or funding.

The next axiom is uncontroversial (its violation is known as the Condorcet paradox):

> **(Axiom 7) Transitivity.** If $T_A \succeq T_B$ and $T_B \succeq T_C$, then $T_A \succeq T_C$.

Finally:

> **(Axiom 8) Independence of Irrelevant Alternatives (IIA).**
> $T_A \succ T_B$ when T_C is present if and only if $T_A \succ T_B$ when T_C is absent.

This requirement invites us to compare two situations, one in which theory T_C is available to the scientist, and one in which it is not. The idea is that the relative ranking of better-evaluated theories should not be affected by the presence or absence of a weaker competitor. Otherwise, it could lead to the following odd situation. A scientist assesses T_A and T_B, and concludes that the best theory is T_B. A colleague then says: "You should also consider T_C." The scientist reassesses the options and concludes that the best theory is T_A. Requiring IIA avoids such reversals and thus seems reasonable in theory choice.

Is there any ranking rule that satisfies these axioms? Transposing a result by Balinski and Laraki (2020, p. 447, Theorem 3), one obtains:

> **Theorem 1.** There exists infinitely many ranking rules that satisfy Axioms 1–8.

The contrast with Arrow's (1951) famous impossibility result is striking: asking scientists to submit grades rather than rankings transforms an impossibility into infinitely many possibilities. One example of suitable rules are *point-summing* methods: each grade is associated with a number, and the numbers received by a theory are averaged according to the weights assigned to the criteria. For instance, 10 points might correspond to *Excellent*, 7 to *Very Good*, 5 to *Good*, and so on. However, these point-summing rules display important drawbacks: the numerical correspondences lack justification and have no meaning beyond stipulation (if they did, the scale would no longer be ordinal, cf. Section 2.1). Moreover, the rule is highly sensitive to errors: because of the averaging

mechanism, a single incorrect assessment on one criterion always affects the outcome, sometimes substantially. This suggests adding another axiom to limit the possible influence of errors.

Suppose that "correct" values can be defined for the grades assigned by the scientist (e.g. those they would assign without bias), and suppose that these values lead to $T_A \succeq T_B$. Consider a criterion that ranks T_B above T_A. If, by mistake, its grade for T_B is decreased, one should not forbid *a priori* that this affects the aggregated result. Otherwise, this could lead to $T_A \succ T_B$ even though no grade actually favors T_A.[3] Conversely, if the grade for T_B is mistakenly increased, one might wish to restrict its aggregated effect. Since the criterion already ranks T_B above T_A, the error merely reinforces this judgment; there may therefore be room to limit its influence. One may thus consider the following requirement: If $T_A \succeq T_B$, then no criterion that grades T_B above T_A should be able to raise T_B's aggregate measure. Symmetrically, it should not be able to lower T_A's group measure. We obtain:

> **(Axiom 9) Error-proofness.** If $T_A \succeq T_B$, then no criterion that grades T_B above T_A can raise T_B's aggregate measure or lower T_A's.

This axiom is also referred to as *strategy-proofness in ranking* in other contexts. When grades are interpreted as agents, it states that agents who disagree with the aggregated ranking cannot benefit from strategically exaggerating their grades. This interpretation can also be defended in the context of theory choice: scientists biased against a theory might strategically exaggerate their assessments on some criterion, providing an additional rationale for Axiom 9. However, the latter turns out to be too demanding (Balinski and Laraki 2020, theorem 5 p. 456):

> **Theorem 2.** No ranking rule satisfies Axioms 1–9.

This motivates considering a weaker version of the axiom. Call two theories polarized when, for all criteria, the higher some criterion evaluates one theory, the lower it evaluates the other. If theories are not polarized, the criteria do not all pull in opposite directions, so the aggregated result may be more robust and the consequences of error (or of strategic manipulation) less dramatic. Hence Axiom 9 is more useful when applied primarily to polarized pairs of theories, rather than to all pairs. An analogue of Theorem 6 by Balinski and Laraki (2020) gives:

[3] This would contradict a Domination axiom, which can be derived from Axioms 1 to 8, cf. Balinski and Laraki (2020, p. 447, Theorem 3).

Theorem 3. There is exactly one ranking rule which satisfies Axioms 1–8, and Axiom 9 on the restricted domain of polarized pairs of theories — call it Majority Judgment.

(Note that Majority Judgment applies to any theories, whether polarized or not.)

2.3 Majority Judgment, the theory choice version

Majority Judgment works as follows (assuming for simplicity equal weights between criteria).[4] For each theory, the grades received for each criterion are ranked in decreasing order, and the median grade becomes the *majority grade* — the overall, aggregate grade for the theory. This is the grade for which a majority of criteria consider that the theory deserves at least this grade, and another majority consider it deserves at most this grade. Theories are then ranked according to their majority grades. In case of a tie, the grades immediately surrounding the median are compared. If these are also equal, the comparison extends to grades farther from the median. When comparing grades away from the median, one theory ranks higher if: (a) its grades are higher, or (b) its grades are closer together (indicating greater consensus). For instance, the astronomer in Section 2.1 assigns grades {*Good*, *Good*, *Fair*} to Ptolemy, and {*Very Good*, *Good*, *Bad*} to Copernicus. The median grade is *Good* is both cases, so one has to look at the first and third grades. *Good* and *Fair* differ by only one grade level, while *Very Good* and *Bad* differ by two; hence, by rule (b), Ptolemy's theory ranks higher. However, if the weight of simplicity is doubled, the grades become {*Good, Good, Fair, Fair*} for Ptolemy, and {*Very Good, Very Good, Good, Bad*} for Copernicus, making Copernicus the winner by rule (a).

Majority Judgment offers a principled way to aggregate qualitative assessments with weights, something Kuhn did not offer. It does more than rank: each theory receives a majority grade. Being the best theory with a majority grade of *Very Good* does not convey the same evaluation as being the best with only *Fair*. By eliciting grades rather than rankings for each criterion, Majority Judgment preserves this richer information to the aggregate level. A key feature of Majority Judgment is that it does not matter which criteria gave which grade, only the set of grades matter. In other words, the algorithm does track the origin of grades (this property is a consequence of the listed Axioms, not an additional or optional feature).

[4] Other expositions of Majority Judgment, some with weights, are Balinski and Laraki (2007), (2011), (2020) for political contexts, Boyer-Kassem (2025) for an epistemic one.

Returning to the objection to the Completeness axiom, the concern in Section 2.2 was that it may not be desirable in specific situations, such as when theories excel on different criteria. A simple, easily implementable solution would be: collect grades as before, use Majority Judgment to provisionally rank theories, but if certain grade profiles appear among the top-ranked theories (to be specified depending on the theoretical objections against the axiom), then veto the Majority Rule ranking, and treat the top two theories (say) as deserving to be equally taught or funded.

3 Theory choice at the community level

So far, we have considered an individual scientist making a choice for themselves. Consider now a scientific community of n scientists — for instance, all astronomers at the turn of the 17th century. Suppose an external observer (e.g. a philosopher or historian) wants to describe the epistemic judgments *of the community* regarding rival theories. Or suppose a central planner wants to allocate research funding according to community views (funding more highly graded theories). In both cases, the community's views on the competing theories must be aggregated. Scientists may not rely on the same criteria of theory choice. Let p now denote the number of distinct criteria within the scientific community. The problem is then to aggregate the $n \times m \times p$ grades into a collective ranking of the theories.[5]

The previous section argued that the $m \times p$ views of an individual should be aggregated using Majority Judgment. In a complementary way, Boyer-Kassem (2025) has argued that the ordinal views of n scientists on a single question (e.g. "How do you evaluate the simplicity of this theory?") should also be aggregated using Majority Judgment. Thus, in the collective and more complex case, Majority Judgment appears to be required twice. However, there are two ways to proceed, depending on what is aggregated first: either the p criteria of theory choice (i), or the n scientists' views (ii).[6]

In procedure (i), called *scientist-based majority judgment*, each scientist assigns an overall grade to each theory by aggregating criteria with Majority Judgment, as in Sect. 2. These overall judgments are then aggregated across scientists using Majority Judgment again. In procedure (ii), called *criterion-based majority judgment*, scientists' grades are first aggregated by criterion using Majority Judgment, yielding the community's majority

[5] If a scientist has not considered a criterion, the weight attributed to it is 0, and the grade may take any value.

[6] These procedures are considered by Balinski and Laraki (2010, chap. 21) in a wine-tasting context.

grade for each criterion.[7] These collective grades are then aggregated across criteria using Majority Judgment (as in Sect. 2, with the community as the agent and normalized weights). One might wonder whether these two procedures are equivalent. They are not, as the counterexample in Table 1 shows — in other words, aggregating with Majority Judgment across criteria and across scientists does not commute. This can be seen as a multicriteria version of the discursive dilemma (Pettit 2001) under Majority Judgment.

		Scientist			
	Criterion	1	2	3	MJ
Theory T_A	α	G	G	F	G
	β	G	F	F	F
	γ	F	G	F	F
	MJ	G	G	F	G or F?
Theory T_B	α	G	G	F	G
	β	G	F	G	G
	γ	F	F	F	F
	MJ	G	F	F	F or G?

Table 1: Example of grades assigned by three scientists to two theories on three criteria (equal weights). G = *Good*, F = *Fair*. Under procedure (i), grades are first aggregated by theory (column) using Majority Judgment, then across scientists (grey cells). Under procedure (ii), grades are first aggregated by criterion (rows), then across criteria. Procedure (i) ranks T_A above T_B, whereas procedure (ii) yields the opposite, showing the procedures are not equivalent.

Which procedure is preferable? Recalling Kuhn's insight that criteria are imprecise and open to interpretation, one faces the question of whether two interpretations of a criterion should count as the same criterion or as different ones. This is not merely a definitional issue, as it can affect the outcome of the criterion-based procedure (ii). For instance, if criteria differ across scientists, Table 1 would contain nine entries for theory T_B, whose grades should be aggregated in a single step using Majority Judgment. The majority grade of T_B would become *Fair*, rather than the *Good* obtained previously.[8] In practice, historians or central planners may have difficulty determining whether criteria are interpreted uniformly within a community. This suggests that the criterion-based

[7] The weights assigned by individual scientists to the criterion must first be normalized. For instance, if two scientists assign weights of 0.2 and 0.3 to coherence, these become 0.4 and 0.6 after normalization, so that they sum to 1.

[8] In the limiting case where all scientists have idiosyncratic interpretations of all criteria, procedure (ii) becomes what Balinski and Laraki (2010, p. 384) call *multicriteria majority judgment*.

procedure, in its non-degenerate form, is viable only when criteria have explicit definitions shared by all. Research project evaluations by external reviewers, as commonly required by funding agencies, typically meet these conditions, whereas historical reconstructions do not. Procedure (ii) therefore seems viable only if one accepts to restrict the epistemic value judgments scientists may express, as funding agencies effectively do. Nonetheless, an advantage of this procedure is that it yields a collective judgment on each criterion, which can help articulate reasons at the collective level. A trade-off thus emerges between freedom in epistemic value judgments and collective justification.

In contrast, scientist-based majority judgment (procedure (i)) accomomodates any degree of heterogeneity in scientists' criteria, without raising difficult boundary questions. It does not require the set, meaning, or weights of the criteria to be externally fixed. Hence, it preserves scientists' epistemic autonomy, is always applicable, and is more robust. However, it does not allow the group to justify its view by appealing to a collective assessment on the criteria themselves. Requiring each scientist to give a single overall grade for each theory (obtained by applying Majority Judgment across criteria) should not be seen as discouraging prior deliberation within the community about how theories perform on individual criteria. Rather, detailed assessments remain relevant at the individual level; only the final aggregation requires each scientist to report an overall judgment. Scientist-based Majority Judgment (procedure (i)) may therefore be regarded as the default procedure.

4 Conclusion

How should scientists rank rival theories? I have proposed a new framework in which scientists assign qualitative grades from a common scale to each theory on each criterion, rather than ranking them. This task is both more informative and typically feasible. At the individual level, the choice of which theory to pursue should be made with Majority Judgment. In funding decisions, one may depart slightly from this rule by allowing some rival theories to remain unranked. At the collective level, for instance when historians reconstruct the views of a scientific community, Majority Judgment should again be used. Two procedures are possible, depending on whether the aggregation first concerns criteria or scientists. Each has its own advantages, though scientist-based Majority Judgment can serve as the default procedure. These results shed new light on the recent debate over whether social choice theory and Arrow's theorem apply to scientific theory choice. By replacing rankings with qualitative grades, we avoid the impossibility results of social choice theory and enable meaningful aggregation using Majority Judgment. In

this sense, Kuhn's "no unique algorithm" thesis is made precise: the present framework provides a formal model that makes sense of it. When scientists evaluate theories using different criteria, weights, and interpretations, Majority Judgment offers a principled way to aggregate their assessments.

References

Arrow, Kenneth J. 1951. *Social Choice and Individual Values.* Yale University Press, New Haven, CT. 2nd edition 1963, Wiley, New York, NY.

Balinski, Michel, and Rida Laraki. 2007. "A theory of measuring, electing and ranking". *Proc Natl Acad Sci USA* 104(21):8720–8725.

Balinski, Michel, and Rida Laraki. 2010. *Majority Judgment: Measuring, Ranking, and Electing.* Cambridge, MA: MIT Press.

Balinski, Michel, and Rida Laraki. 2020 "Majority Judgment *vs* Majority Rule", *Social Choice and Welfare*, 54: 429–461.

Boyer-Kassem, Thomas. 2025. "Majority Judgment for Collective Beliefs and Science", working paper. `https://philarchive.org/rec/BOYMJF`

Dietrich, Franz, and Christian List. 2016. "Probabilistic Opinion Pooling". In A. Hájek and C. Hitchcock, eds, *Oxford Handbook of Probability and Philosophy*, Oxford University Press.

Kuhn, Thomas S. 1977. Objectivity, Value Judgment, and Theory Choice. In T. S. Kuhn, *The Essential Tension*, Chicago: University of Chicago Press, p. 320–339.

Morreau, Michael. 2015. "Theory Choice and Social Choice: Kuhn Vindicated". *Mind* 124(493), 239–262.

Okasha, Samir. 2011. Theory choice and social choice: Kuhn versus Arrow. *Mind* 477, 83–115.

Okasha, Samir. 2015. On Arrow's theorem and scientific rationality: reply to Morreau and Stegenga. *Mind* 493, 279–294.

Pettit, Philip. 2001. Deliberative democracy and the discursive dilemma. *Philosophical Issues*, 11, 268–299.