Historical and Theoretical Insights for Coarse-Grained Evaluation in Theory of Justice

*Takashi Izumo**

Abstract

The maxim "suum cuique" (to each their due) involves assessing person-related features, but institutions rarely achieve precise measurement without errors or costs. They therefore depend on coarse-grained evaluation, which intentionally reduces evaluative resolution, and treat differences within a tolerated band as practically equivalent. A common form is categorical evaluation, in which continuous values are divided into a few status-relevant categories, potentially resulting in cliff-edge effects at the boundaries. While natural sciences have developed systematic methods to model and control error, normative inquiry has discussed analogous forms of simplification less systematically and without a unified framework. This paper offers a justice-based perspective on categorical evaluation and explores the normative issues it raises. It draws on historical insights from Aristotle's equity, Jesuit probabilism, Thomasius's cognitive pessimism, and Kant's regulative goals. By comparing cases from grading and threshold-based fiscal and eligibility rules, the paper argues that how banding should be assessed is not determined by banding itself. Instead, it depends on which evaluative factors are prioritized, such as efficiency and ease of administration, transparency, incentive effects, and the availability of boundary relief.

1. Introduction

On justice, Ulpianus, one of the most influential Roman jurists of the classical period, famously provided the following formulation preserved at the beginning of Justinianus's *Institutiones*: "Justice is the constant perpetual will to render to each their due" (Iustitia est constans et perpetua voluntas ius suum cuique tribuens).[1] "To each their due" can be interpreted as a directive to assign claims and burdens based on person-related features.

* Takashi Izumo, Dr. Dr. jur. Professor at Nihon University College of Law, Tokyo, Japan. Email: izumo.takashi@nihon-u.ac.jp

1) Behrends, O. et al. (Hrsg.). (1997). *Corpus Iuris Civilis: Text und Übersetzung: Institutionen*, 2. Aufl., Heidelberg, C. F. Müller. S. 1.

However, before debating which features justify desert or entitlement, a fundamental formal challenge arises: no institution can distinguish arbitrarily fine differences at scale. In practice, institutions must coarse-grain what is owed: they define ranges and treat differences within those ranges as practically irrelevant for assessment and distribution.[2] University letter grades illustrate this: a 75 and a 73 both receive a C, while a 70 and a 69 can fall on opposite sides of a boundary (C vs. D).

In this paper, the author adopts coarse-grained evaluation as a central theme in normative inquiry. By coarse-grained evaluation, the author refers to institutional designs or assessment procedures that intentionally reduce evaluative resolution, treating some differences among cases as practically negligible for purposes such as assessment, allocation, or responsibility. Coarse-graining can take various forms. It may involve numerical approximation (e.g., rounding or quantization) or the use of intervals, bands, or other simplified representations. The paper emphasizes a prominent form of coarse-graining: categorization. By categorical evaluation, the author means the practice of mapping a continuous range of values onto a limited number of status-relevant classes or bands and treating differences within a class as equivalent for institutional purposes. In this context, categorization is not just about using a representative number for convenience. Assigning a representative value to a band is a separate process that can facilitate coarse-graining. However, the core feature of categorization is the assignment of status itself, especially when it determines eligibility, duties, or other qualitative outcomes. For example, when raising taxes on high-income earners, the externally conferred status of "high-income earner" itself plays a critical role.

This paper aims to provide a foundational inquiry into the role that categorical evaluation plays in theories of justice, outlined as follows. Section 2 formulates the problem of categorical evaluation and connects it to related themes in moral and legal thought. Section 3 compares the way errors are handled in the natural sciences with their treatment in normative inquiry, emphasizing the differences. Section 4 analyzes threshold-based eligibility rules in Japan's social insurance schemes and uses them to explore the cliff-edge phenomenon, characterized by sudden changes in status at a bound-

2) Izumo, T., and Weng, Y.-H. (2022). Coarse ethics: how to ethically assess explainable artificial intelligence. *AI and Ethics* 2(3):449–461. https://doi.org/10.1007/s43681-021-00091-y; Izumo, T. (2025). Introduction to coarse ethics: Tradeoff between the accuracy and interpretability of explainable artificial intelligence. In: Montag, C., and Ali, R. *The Impact of Artificial Intelligence on Societies.* Studies in Neuroscience, Psychology and Behavioral Economics. 155–167. https://doi.org/10.1007/978-3-031-70355-3_12

ary. Section 5 offers a conclusion.

2. Problem Definition

The maxim "render to each their due" (in Latin: suum cuique tribuere) can be understood, in legal terms, as a directive to assign each person a correlative right or duty based on some person-relative feature. Specifically, if A and B collect 90 nuts together and their contributions are judged to be in a 2:1 ratio, then giving 60 to A and 30 to B might seem to be just. I say might advisedly. What counts as the relevant feature—such as effort, time spent, marginal productivity, need, prior claim, risk borne, or a combination—does not belong to formal justice but to substantive justice, and there is no universally accepted criterion for it.

This paper does not try to define the reasons for substantive justice. In the nut-division example, the question of what should be considered is not our focus. Our starting point is the idea that even at the level of formal justice, that is, before deciding what each person's due is, errors can occur because of how legal and administrative systems must carry out directives. Just for the sake of argument, assume that time spent is used as a substantive basis. This assumption is controversial and may be hard to defend in practice; some might argue that working hours should not serve as a basis for distribution, but that debate is not our concern here. Suppose A and B agree to split the nuts based on the time spent. A suggests: "I worked for 1 hour while you worked for 30 minutes; let's split 2:1." B responds: "You worked 59 minutes and 37 seconds, and I worked 30 minutes and 1 second; under that agreement, a strict 2:1 split isn't justified." How should we interpret such precise descriptions? In everyday life, many consider it nonsense. If two people agree to split a bill because each ate one pizza of the same kind and size, it would be silly for one to object: "Your pizza weighed 156 grams while mine was 155 grams; splitting evenly is unfair." Even if the gram difference were proven, the complaint would still seem out of place.

Let's compare this issue with various schools of legal thought to identify its position within the field of ethics.

(1) Aristotle's Equity (επιείκεια)

Aristotle recognized that rules written in natural language are ill-suited to capture subtle differences. Law must "speak in universals," and because of this universality, it can fail at the edges.[3)] Equity (επιείκεια) is the tool for

3) Aristotle, *Nicomachean Ethics*. Translated by W. D. Ross. NetLibrary, Inc. Vol. 5. §. 10., p. 58.

correcting a rule after the fact when, without such correction, applying its general wording as written would otherwise yield an unfair result—one that the legislator would have avoided by adding an explicit exception or proviso had they been present and known the facts (NE 5.10, 1137b11–1138a3).

Put simply, Thomas Aquinas (ca. 1225–1274) uses the well-known "city gates" example to explain why a generally sound rule might need an exception: a rule like "Whoever opens the city gates shall suffer death" can be reasonable in normal times (since it prevents treachery), but during a siege, the same act—opening the gates to fight off an attack or to bring in help—should not result in the penalty.[4)]

Aristotle's account of epieikeia aligns with our issue as it diagnoses failures in rule application. However, it addresses a different question. Epieikeia asks how to correct an otherwise reasonable general rule after the fact when it fails in a specific case; it is a theory of post-hoc correction. In contrast, this paper examines why, and under what conditions, a legal or moral framework can justify ignoring small differences beforehand, meaning to embed indifference to minor variations directly into the rule.

(2) Catholic Probabilism

A second related concept is probabilism in early modern moral theology, often connected with Jesuit casuistry. Probabilism asserts that if there is a probable opinion supporting the permissibility of an action (in Latin: opinio probabilis), an individual can legitimately follow it even if the opposite opinion (impermissibility) is more probable.[5)] Its classic formulation is attributed to Bartolomé de Medina (1527–1580): "If an opinion is probable, it may be followed, even if the opposite opinion is more probable" (in Latin: Si est opinio probabilis, licitum est eam sequi, licet opposita probabilior sit).[6)]

Probabilism involves overlooking differences when it allows equally acceptable choices among competing views despite their different likelihoods. However, it does not fully address our issue. Our concern is not that current evidence is insufficient to determine which theory of justice is better, thus allowing us to choose any theory that is sufficiently probable. Instead, while holding a substantive basis constant, we ask when a rule

4) Aquinas, T. (1892). *Opera Omnia: Prima Secundae Summae Theologiae*. Tom. 9. Romae: Typographia Polyglotta. q. 96. a. 6. p. 187.
5) Franklin, J. (2015). *The Science of Conjecture: Evidence and Probability Before Pascal*. Baltimore, Johns Hopkins University Press. pp. 74–76.
6) Schüssler, R. (2005). On the Anatomy of Probabilism. In: Kraye, J., and Saarinen, R. (eds.). *Moral Philosophy on the Threshold of Modernity*. 91-113. Cham, Springer. p. 92.

can preemptively treat micro-differences as irrelevant from a normative perspective, meaning it intentionally incorporates indifference into the rule itself even when all facts are known.

(3) Thomasius on Folly

Christian Thomasius (1655–1728), skeptical of human insight, advocated practical maxims like the psychological golden rule—"do not do to others what you would not have done to you"—rather than complex casuistry.[7] Thomasius believed that ordinary citizens could not engage in legal reasoning from the start and thus could not make judgments based on it.[8]

In spirit, this relates to error handling: if subtle differences exceed ordinary understanding, we should not design systems that rely on their detection. However, our thesis is not about pessimism of human nature ("humans are foolish, therefore we should accept rough rules"). We argue something more precise and different: even when fine distinctions can be identified, institutions may be normatively justified in ignoring them for allocation purposes because striving for precision can distort incentives, hinder publicity and coordination, and create unfairness at boundaries.

(4) Kant's Regulative Ideal

Immanuel Kant (1724–1804) argues that reason sets ideals in *Critique of Pure Reason*, meaning strict and complete notions of virtue or legality, which serve as regulative guides rather than constitutive rules for human action (A569/B597).[9] These ideals help shape judgment and promote self-improvement, but they are not standards we are required or even able to meet precisely in practice. His common example is the Stoic sage: an ideal of moral perfection that acts as a guiding star, though no one is expected (or able) to become such a sage (A569/B597).[10] In the *Metaphysics of Morals*, he also criticizes types of moral fanaticism or pedantry: an overly zealous obsession with moral details that confuses the idea with a decision procedure, thereby distorting normal deliberation (MS TL §17, AA 06: 409.13–19).[11]

Among the four comparanda, Kant is the closest to our target. His ex-

7) Thomasius, C. (1718). *Fundamenta juris naturae et gentium*. Halle, Christoph Salfeld. Lib. 1. Cap. 6. §. 42., p. 177.
8) *Ibid*., Lib. 1. Cap. 6. §§. 19–22., pp. 171–172.
9) Kant, I. (1996). *Critique of Pure Reason*. Translated by Werner S. Pluhar. Cambridge, Hackett Publishing Company. p. 561.
10) *Ibid*., p. 562.
11) Kant, I. (2017). *The Metaphysics of Morals. Revised edition*. Edited by Lara Denis. Translated by Mary Gregor. Cambridge, Cambridge University Press. p. 178.

planation of regulative ideals and flexibility in practical judgment suggests that when multiple agents follow a common public rule, conformity to that rule need not—and often should not—depend on tracking small differences. The key question is: how did Kant suggest we judge the acceptable error when finite agents operate under the same rule? It appears he did not provide a detailed explanation. He states that the ideal is regulative, not a strict blueprint, but gives little detail on how regulation works, such as what level of latitude is justified, how much variance is acceptable, and who bears the residual error. It's like saying, "a product should conform to the blueprint, but tolerances are allowed," without explaining how to determine those tolerances.

3. Comparison Between Natural Science and Normative Science

Why, even though the problem of error was already recognized by the late 18th century, has it still not received systematic treatment up to today? The reason becomes clear when compared to the natural sciences.

By the mid-17th century, natural philosophers began to treat experimental error as a distinct epistemic issue. Influenced by Baconian critiques of "idols," figures like Robert Boyle (1627–1691) and Robert Hooke (1635–1703) specifically studied and listed "popular errors," arguing that careful methods could uncover and fix them.[12] By the early 18th century, this empirical approach was formalized mathematically: Thomas Simpson (1710–1761)'s 1755 memoir used probability axioms in telescopic observations, treating positive and negative errors as equally likely.[13] In the late 18th century, Pierre-Simon Laplace (1749–1827) produced foundational work in probability, including influential memoirs beginning in 1774, and he contributed to the emerging mathematical treatment of observational uncertainty.[14] In short, early modern science promoted a systematic approach to error analysis; measurement uncertainty was recognized, averaged, and modeled (in astronomy, physics, and other fields) as a key part of knowledge, culminating in the development of least-squares and probabilistic theories by the end of the century.

By contrast, mainstream moral philosophy during this period remained firmly rooted in rationalism and rule-based approaches, with little explicit

12) Leonard, A., and Parker, S.E. (2022). Put a mark on the errors: Seventeenth-century medicine and science. *History of Science* 61(3):287–307, pp. 303–305. https://doi.org/10.1177/00732753221135046

13) Shoesmith, E. (1985). Thomas Simpson and the arithmetic mean. *Historia Mathematica* 12(4):352–355, p. 352. https://doi.org/10.1016/0315-0860(85)90044-8

14) Wilson, E.B. (1923). First and second laws of error. *Journal of the American Statistical Association* 18(143):841–851, p. 841. https://doi.org/10.2307/2965467

acknowledgment of uncertainty. Hugo Grotius (1583–1645), Samuel von Pufendorf (1632–1694), Christian Wolff (1679–1754), and their followers saw natural law as a deduction from right reason, meaning they believed moral law was dictated by right reason.[15] Kant later summarizes this tradition by stating as follows.

> The practical rule is therefore unconditional and so is represented a priori as a categorical practical proposition by which the will is objectively determined absolutely and immediately (by the practical rule itself, which accordingly is here a law). For, pure reason, *practical of itself*, is here immediately lawgiving.[16] (KpV, AA 05: 31.7–10)

It must be noted that Grotius and other natural law theorists were not indifferent to the gap between law and reality, nor to the issue of factual determination in judicial proceedings. Grotius acknowledged that overly detailed natural law rules could lead to excessive litigation. For example, he argued that there should be a precise equivalence in value between a good and its price. If the amount paid is more or less than the intrinsic worth of the item, moral duty requires correcting the discrepancy.[17] However, he also recognized that allowing legal remedies for every minor imbalance would result in an unmanageable surge of lawsuits. Therefore, he concluded that such cases should be considered morally significant but not legally actionable.[18] Pufendorf offered a partial revision by allowing legal recourse in cases where the imbalance between a good and its price is "graver" (in Latin: gravior).[19] However, he did not specify clear criteria for determining what qualifies as a "graver" disparity. Furthermore, as mentioned earlier, Kant introduced the concept of the regulative function of ideals, thereby allowing a distinction between the ideal and the empirical realms. Although

15) Grotius, H. (1680). *De jure belli ac pacis.* Amsteldam, Janssonio-Waesbergios. Lib. 1. Cap. 1. §. 10., p. 6.: "Jus naturale est dictatum rectae rationis"; Pufendorf, S. (1672). *De jure naturae et gentium*. Lund, Adam Junhans. Lib. 2. Cap. 3. §. 13., p. 179.: "Igitur hoc sensu lex naturalis nobis dictamen rectae rationis asseritur, quod intellectui humano ea sit facultas, ut ex contemplatione conditionis humanae liquido perspicere possit, ad norman ejus legis sibi necessario vivendum: simulque investigare principium, ex quo ejusdem praecepta solide et plane demonstrari queant."; Wolff, C. (1754). *Institutiones juris naturae et gentium*. Halle, Officina Rengeriana. Part. 1. Cap. 2. §. 39., pp. 20–21.: "Vocatur autem naturalis, quae rationem sufficientem in ipsa hominis rerumque essentia habet: [...] Lex naturae communiter quoque Jus naturae appellatur."

16) Kant, I. (2015). *Critique of Practical Reason*. Translated by Mary Gregor. Cambridge, Cambridge University Press. p. 28.

17) Grotius, *supra note* 15, Lib. 2. Cap. 12. §. 11. n. 1., p. 255.

18) *Ibid*., Lib. 2. Cap. 12. §. 26. n. 1., p. 264.

19) Pufendorf, *supra note* 15, Lib. 5. Cap. 3. §. 9., pp. 629–630.

these positions resemble the situation of a discrepancy between theory and observation in the natural sciences, the natural law theorists, unlike natural scientists who relied on probability theory and statistics, did not believe that such a discrepancy could be resolved through mathematical means.

This paper does not argue that ethics or jurisprudence should incorporate probability theory or statistics into rule-setting in the same way as the natural sciences. Such an application, in any direct sense, is impossible. This inapplicability stems from a fundamental difference between the natural sciences and the normative sciences. In the natural sciences, it is the theory that is subject to verification through the relation between theory and observation or experiment, and thus it is accepted or rejected; the observation or experiment itself is not directly verified. For example, Newton claimed that chromatic aberration—a distortion caused by the dispersion of light—could not be eliminated even by combining lenses made of different materials; however, this claim was later disproven by the invention of the achromatic lens.[20] In this case, what was ultimately rejected was Newton's theoretical assertion, not the empirical findings demonstrated by the achromatic lens. In contrast, in the normative sciences, the relationship between theory and reality is such that reality itself is evaluated and either endorsed or condemned. When a natural-law rule states that "one must not kill," but an act of killing still happens, it is not the rule that is refuted; rather, the act of killing is judged wrongful. This does not mean that normative theories are never revised or rejected. They can be refined, qualified, or abandoned based on feasibility constraints, empirical consequences, and reflective equilibrium. For example, if doctrines in criminal law were based on an incorrect physiognomy claiming to identify criminal tendencies from facial features, those doctrines would be expected to be revised once this so-called "science" is shown to be methodologically flawed.

Accordingly, the way errors are treated in the natural and normative sciences initially differs in the types of questions they ask.

- Natural sciences: When testing a theory (an ideal model), how much can a theoretical prediction differ from the observed result and still be considered acceptable?
- Normative sciences: When assessing an action, how far can it deviate from the norm (the ideal model) and still be considered acceptable?

20) Glashow, S.L. (2008). The errors and animadversions of Honest Isaac Newton. *Contributions to Science* 4(1):105–110. p. 108.

4. Case Studies

(1) Letter Grading

This paper does not aim to present a definitive solution to the previous asymmetry but instead to enhance our understanding through a series of illustrative cases. The first case involves a situation where a very detailed norm proves too challenging to implement in practice and is thus replaced by a less precise one. Consider the following example: Suppose that, when grading a jurisprudence report, we use a 101-point scale ranging from 0 to 100. We also assume the normative principle that "students who perform better should receive higher evaluations." Although this assumption is open to debate, it is reasonable within normal limits. For example, claiming that a student who earns 40 points has written a better report than one who earns 30, or that a student with 70 points has produced a better report than one with 65, seems convincing—especially when compared to more controversial moral judgments, such as whether volunteering or donating money results in the greater good.

In actual university grading systems, however, such a 101-point scale is often converted into a single-letter grade scale. For example, scores below 60 are marked as F; scores from 60 to 69 are marked as D; scores from 70 to 79 are marked as C; scores from 80 to 89 are marked as B; and scores from 90 to 100 are marked as A, creating a five-tier system. Additionally, this may be simplified further into a binary distinction, with F representing failure and A, B, C, and D representing passing. This exemplifies what this paper refers to as setting an admissible margin of error. In such a five-tier system, it is implicitly assumed that the difference between, for instance, a student who scores 65 and one who scores 67 can be legitimately disregarded. Therefore, even if Student s_1 claims, "I scored 75 while s_2 scored 73, so it is only just that I be rated higher," the instructor can respond, "According to the rules of this examination, everyone scoring between 70 and 79 receives the same grade of C; hence, the evaluation is not unjust."

Therefore, instead of using exact values in assessment, we will refer to the method utilizing fixed ranges as "categorical evaluation." All the following are examples of categorical evaluation.

- Treating a person who worked 7 hours and 56 minutes the same as someone who worked exactly 8 hours by calling it "8 hours of work" and paying them the same wage.
- Not differentiating between stabbing someone 5.0 cm and 5.1 cm with a knife, but treating both as injuries of the same severity and applying the same penalty.
- Recognizing both a person who donated 100 million yen and another

who donated 110 million yen as major donors, and awarding both with the same recognition.

In educational research, the validity of such categorical (banded) evaluation has often been a direct object of study. When there are multiple graders, disagreements about rank ordering are common.[21] For example, Grader g_1 might score $s_1 = 70$ and $s_2 = 71$, while Grader g_2 scores $s_1 = 72$ and $s_2 = 70$; they do not agree on who is better. Banded letter grades help make the results more stable: in this example, both s_1 and s_2 fall into the C band, so the assigned grade is the same. Additionally, in microeconomics research, administrative simplicity and clear communication are important: coarse public cut-scores (e.g., A/B/C or pass/fail) make information easier for audiences to process and can even reduce overall estimation error by pooling cases in the middle.[22]

(2) Eligibility Thresholds in Social Insurance

A widely discussed recent example in Japan is the so-called "1.30-million-yen wall." This term was used in debates over social insurance dependency rules. In simple terms, when a spouse's expected annual income is estimated to exceed a set threshold (often around 1.30 million yen), the spouse may lose dependent status under the insured person's coverage. This can result in the spouse having to pay premiums on their own. In policy debates, the threshold is often linked to an exemption system where spouses considered "dependents" below that limit can avoid contributing to pension and health insurance schemes. The arrangement is frequently seen as a disincentive to working longer hours.[23]

What is striking is that this "wall," unlike letter grading in education, which often receives positive appraisal, has tended to be evaluated negatively. Where does this difference come from? In my view, it does not stem from a disciplinary split between education and law per se, but from a difference in focus: educational research often foregrounds the efficiency of grading from the assessor's perspective, whereas theories of justice emphasize the fairness of the institution. If we swap these focal points, negative

21) Schinske, J., and Tanner, K. (2014). Teaching more by grading less (or differently). *CBE—Life Sciences Education* 13:159–166, p. 160. https://doi.org/10.1187/cbe.CBE-14-03-0054

22) Harbaugh, R., and Rasmusen, E. (2018). Coarse grades: Informing the public by withholding information. *American Economic Journal: Microeconomics* 10(1):210–235, p. 211. https://doi.org/10.1257/mic.20130078

23) Makiko Yamazaki, "Japan's labour crunch forces rethink on traditional homemakers," *Reuters*, June 12, 2025, accessed October 23, 2025, https://www.reuters.com/business/world-at-work/japans-labour-crunch-forces-rethink-traditional-homemakers-2025-06-12/

assessments could arise in education, and positive assessments could arise in justice theory. That is, if educational studies were to attend not to reduce the teacher's workload but to fairness among students, and if justice theory were to attend not to fairness among taxpayers but to administrative ease on the taxing side, we could articulate the following:

- Letter grading can trigger perceptions of unfairness among students. Although the numerical gap between a student with 69 points and one with 70 points is only one point, the former receives a D while the latter receives a C, creating a sharp discontinuity. Conversely, unfairness may also be felt between a student with 60 points and one with 69 points: despite a nine-point gap, both fall into the same D category and thus receive the same grade.
- In social security systems, there is a rational cost-based case for adjusting contributions or benefit eligibility using income bands. Creating broad categories—e.g., "dependent" versus "required to enroll/pay contributions as an insured person"—can substantially reduce administrative and compliance costs compared to continuously tracking and recalculating status based on income down to the one-yen level.

From the above, we can conclude that categorical evaluation generally leads to boundary discontinuities, and these discontinuities can be analyzed in various ways. The discontinuity itself has no inherent legitimacy or illegitimacy. Therefore, when rule-makers implement a cliff-edge design, it cannot be justified or criticized without considering the broader context. An additional evaluative perspective is necessary. Although we have considered perspectives such as efficiency, administrative costs, and the psychology of affected parties, it is uncommon, if not nonexistent, in the current literature to systematically treat these perspectives as separate objects of analysis within theories of justice.

5. Conclusion

This paper argues that even in the formal stages of developing theories of justice, creating tolerance bands presents a normative challenge, and that categorical evaluation often results in boundary discontinuities. The justice principle "to each their due" requires us to evaluate each individual's deservingness; however, accurately measuring this deservingness is unrealistic, and errors are frequent. In everyday practice, we accept these errors by categorizing cases into broad groups, treating unlike cases as if they were the same. University letter-grade evaluations exemplify this approach.

We refer to this as coarse-grained evaluation, which intentionally reduces the evaluative resolution. We then define categorical evaluation as a form of such practices in which assigning cases to discrete categories or status bands, rather than merely numerical approximation, is normatively significant.

This issue has been debated since ancient times in various forms that don't necessarily align with the focus of this paper. Aristotle, recognizing the excesses of natural language, permitted corrections after the fact through the principle of equity. Jesuit probabilism stated that when multiple plausible arguments exist, any of them could be chosen. Early modern German jurist Thomasius, modest about human intellectual abilities, recommended intuitive judgment guided by the Golden Rule—"do not do to others what you would not have done to you"—rather than strict reasoning. Kant proposed governing reality by the ideal, allowing a gap between the ideal and actual conditions. Of these, Kant's view most closely relates to this paper's concerns; however, he did not specify how to set the tolerance between the ideal and reality. This suggests that, even when the gap is acknowledged, the question of how to define tolerances remains insufficiently explored in normative theory. Whereas the natural sciences have developed systematic tools for managing error, normative inquiry has addressed analogous issues less systematically and without a widely shared framework.

Categorical evaluation gives rise to several potential problems; among them is the discontinuity at category boundaries. Crucially, this discontinuity is neither inherently good nor bad; its evaluation depends on the context. For example, in education studies, some researchers view letter grades positively as a cliff-type (notch) design. In contrast, in the social security system, such cliff-type designs have faced significant criticism. This does not mean that either judgment is wrong; rather, opinions differ depending on which factors—such as efficiency or fairness—are emphasized. In other words, no inherent justice or injustice exists in the cliff-type design itself. However, there is still no systematic discussion, for instance, on how to compare or balance these factors; therefore, there is a need to develop coarse-grained evaluation as a distinct area within theories of justice.

We have not, in this paper, examined which principles of justice might apply to the issue at hand, nor whether a mathematical model, similar to those used in the natural sciences, could be developed. We also have not explored the historical and philosophical question of whether Kant's idea of the regulative role of ideas could provide insights into resolving, for example, the statement "ought implies can." If "ought implies can" is true, then it may follow that evaluators cannot be reasonably expected to manage detailed distinctions that are operationally unfeasible. These remain tasks

for future research.