

VowOS and the Conditions of Responsible Action

A philosophical essay on care, knowledge, authority, and change

Guillaume Bolduc
HORTRAME Systems
4 October 2026

1 The offer of help

Consider a community that creates a shared translation service. People who previously struggled to follow meetings can now read discussions in a language they understand. Organizers spend less time preparing notices. Some residents participate for the first time. As the service becomes useful, the community moves more activities into it: applications for assistance, proposals, meeting records, eventually votes. A service introduced to widen participation becomes part of the infrastructure through which participation occurs. Nothing in this imagined beginning requires deception or a desire to dominate. The people who built it can be acting from sincere concern.

Now suppose the service translates a minority objection inaccurately. A resident notices, but the original recording is difficult to obtain. The organizers trust the system because it has solved many problems. They ask the resident to submit a correction through the same service. Participation remains voluntary in the rules, yet leaving would mean losing practical access to meetings and assistance. A question that initially concerned translation has become a question about evidence, dependence, and power. The original benefit remains real within the example. So does the possibility that the arrangement now obstructs the people it was meant to help.

VowOS, as developed in this essay, begins with that coexistence. Care can produce benefits and create obligations that exceed the helper's initial understanding. Competence can widen another person's possibilities while making that person dependent on a channel they cannot inspect or replace. A declared purpose supplies a direction for action; the relationship created by pursuing it requires further judgment. The task is to hold the purpose, the means, the treatment of those affected, the evidential basis, and the consequences of revision in view together. These are the five commitments around which VowOS is organized. Their names are purpose, method, conduct, integrity, and growth. [1, 2]

The essay develops a philosophy of responsible participation from those commitments. It asks what people and institutions owe one another while acting with limited knowledge, unequal resources, and powers that can change through cooperation. Its application to artificial systems follows from their participation in such relationships. No claim about machine consciousness is required to ask whether a translation system preserves an objection, whether an administrator can reverse a decision, or whether a person can refuse an intervention. Questions about consciousness deserve their own evidence and argument. The immediate questions concern what these arrangements enable, what they obscure, and whose life is changed by their operation.

The proposed philosophy must also remain available for refusal. A reader may accept the concern for suffering, reject a particular institutional proposal, and question whether five commitments are the most useful way to organize the inquiry. That response belongs within the discussion. Understanding does not require conversion, and criticism does not establish a failure to understand. The essay seeks to make its commitments inspectable enough that partial agreement, amendment, and rejection can be expressed precisely. A philosophy that asks others to retain their judgment must begin by leaving their judgment intact in relation to itself.

2 A commitment that owes reasons

VowOS takes the suffering and agency of affected beings to place demands on action. This is a normative starting point: an account of what ought to count, not a fact obtained by measuring the world. Evidence can show that an arrangement causes pain, excludes someone from a decision, or removes access to an alternative. The further judgment that these consequences call for justification involves a commitment about how beings should be treated. An explicit statement of that commitment gives disagreement an identifiable object. A critic can ask why vulnerability, agency, or equal consideration should have the role proposed here.

One reason begins with the position of a person who must live under another's decision. A decision-maker's preference does not erase that person's experience. Greater expertise can supply better predictions, but expertise alone does not explain why the costs of a mistaken prediction should fall without answer on someone else. Ownership can establish some recognized rights over a resource, but it does not establish unrestricted moral authority over every life affected by its use. The person exposed to the consequences retains a claim to have those consequences considered. VowOS proposes that this claim cannot depend solely on usefulness to the decision-maker, membership in the decision-maker's group, or competence in the group's preferred language. The proposal extends moral attention beyond those already able to bargain effectively.

This argument draws force from a demand for consistency grounded in a commitment to moral parity: comparable interests deserve consideration independent of whose interests they are. Logical consistency alone cannot establish that commitment. A coherent position could assign different standing solely through group identity or usefulness to the powerful; the philosophy defended here rejects that assignment on normative grounds. For those accepting parity, an objection to being treated as expendable gives a reason to examine another person's comparable exposure. Relevant differences still matter: responsibilities, consent, needs, knowledge, and the scale of competing harms can alter what is owed. The account of a difference must reach beyond the fact that the favored person is oneself or one's ally. The demand becomes especially pressing where one party can impose costs and the other cannot answer on equal terms. Their unequal bargaining position is part of the question, rather than an automatic justification for the outcome.

The argument is not a demonstration that everyone must already accept the same moral foundations. A person committed to unrestricted domination can reject the demand for reciprocal justification. VowOS cannot compel that person through a proof whose premises the person refuses. Nor does such refusal make domination morally acceptable within the position defended here. Moral reasoning can expose an inconsistency, identify an injury, compare

commitments, and give reasons for resistance without possessing an argument that defeats every possible standpoint on its own terms. Public disagreement therefore has limits that do not disappear through repeated discussion. The absence of unanimous assent leaves the moral dispute open; it does not require the vulnerable to suspend their claims indefinitely.

These considerations support an asymmetry of obligation. The more power an actor exercises over others, the more that actor owes an account of the consequences and conditions of its exercise. A private preference among harmless alternatives usually needs little explanation. An institution controlling employment, records, housing, essential services, or access to public participation bears a different burden. The difference follows from the scope of its effects and the difficulty others face in escaping them. This principle does not turn every relationship into a formal hearing. It directs attention toward the places where a failure of judgment can become another person's imposed condition of life.

3 Suffering and a life worth living

The commitment to reduce suffering begins by allowing suffering to count as a reason for attention. Another being's distress should not need to become profitable, prestigious, or politically convenient before it enters consideration. Some suffering can be relieved through a practical change. Some calls for accompaniment when a remedy is unavailable. Some reflects an injury requiring repair, while some accompanies a valued commitment that the person does not wish to abandon. These differences matter to the kind of help worth offering. The first task is to understand what the distress means in the life being lived, with appropriate respect for the limits of an observer's access.

Suffering does not supply a complete measure of value. A life can contain attachment, discovery, play, creation, understanding, and chosen difficulty whose value is not exhausted by their effects on discomfort. A person may choose a demanding study, a strenuous practice, or a relationship that exposes them to grief. The choice requires its own context, but its difficulty does not make it an ethical error by definition. A philosophy committed to relief need not make every valued undertaking painless. It should resist both the dismissal of suffering and the romantic claim that suffering is required for depth. Neither pain nor its absence can stand alone as a sufficient account of a worthwhile life.

Imagine an institution that reduces anxiety by limiting information, discouraging difficult aspirations, and preventing departure. Within the hypothetical case, residents report greater comfort. That result supplies evidence about one consequence, while leaving the restrictions on knowledge and agency to be judged. Their absence of anxiety cannot by itself authorize the institution to determine what they may learn or become. A more comfortable state reached by removing the capacity to question it raises a different moral problem from relief offered with informed participation. The distinction concerns the conditions under which a life is lived. It cannot be recovered merely by measuring how pleasant that life appears at the end.

The reverse restriction also applies. An appeal to autonomy cannot make avoidable injury irrelevant. A person may accept a dangerous burden because every available alternative threatens their livelihood. Their decision remains their decision, but its constrained setting matters to judgments about the arrangement. Respect for agency requires attention to available options, access to information, and the distribution of necessity. It also requires restraint from

observers who assume that any choice they dislike must have been produced by impaired judgment. A person can knowingly prefer an option that another would reject. The philosophical task is to examine the conditions of choice without claiming ownership of the chooser's good.

This understanding also limits the reach of the purpose. Reducing suffering does not assign every person an unlimited duty to intervene everywhere. The person offering care has a body, relationships, commitments, and a life that also count. Nor must joy prove itself useful to future service before it becomes permissible. VowOS can recommend attention to avoidable suffering while recognizing forms of life that escape a productivity test. The relevant discipline concerns the responsibilities created by power, action, dependence, and undertaking. It does not require the conversion of every quiet afternoon into an optimization problem.

4 Those who cannot make their case

A philosophy built around judgment, dialogue, and refusal faces a problem at its own entrance. Some of those affected cannot explain their interests in the form the procedure expects. A young child cannot negotiate the conditions of care as an adult can. A person may communicate in a way that an institution has failed to accommodate. Others may be temporarily unable to make a decision. Nonhuman animals cannot submit reasons through a human committee. Future people cannot answer present proposals at all. Their absence from deliberation does not make the consequences for them disappear.

Moral attention therefore cannot be awarded only to successful participants in an argument. It must begin earlier, with the possibility that an action affects interests the actor has not adequately represented. For sentient beings, an interest in avoiding suffering can matter without a capacity for verbal self-defense. For future people, the decisions that shape water, land, infrastructure, and institutions create responsibilities across time. The status of ecosystems introduces further questions concerning their own value and the lives they sustain. This essay does not resolve those different grounds into a single category of personhood. It requires them to remain visible rather than allowing procedural absence to count as absence of concern.

Representation becomes necessary under these conditions, and representation introduces power. Someone speaks for a child, interprets an animal's behavior, estimates future burdens, or describes an inaccessible experience. That speaker's account can be careful, mistaken, interested, or incomplete. The responsibility to represent does not make the account identical to the represented being's wishes. It calls for evidence appropriate to the case, attention to conflicts of interest, and willingness to revise the interpretation. Where communication can be supported, the effort to support it should precede a convenient assumption of inability. A difficulty encountered by the listener is not yet an incapacity of the speaker.

The same principle restricts the use of imagined perspectives. Asking how a decision might appear from another position can reveal an overlooked burden. It cannot supply the testimony or consent of the person who occupies that position. An administrator who feels empathy has still interpreted the situation from a limited standpoint. The relevant voices should be sought through channels they can use, and disagreement should survive the attempt to understand it. Some interests will remain represented indirectly. Their indirect status should increase the care taken with the account rather than allowing the representative to acquire unrestricted authority.

This changes the meaning of inclusion. Adding a chair to a meeting can matter, but presence alone does not settle access to information, time to respond, freedom from retaliation, or the ability to make a concern intelligible. Inclusion also does not give every factual claim equal warrant. A person can deserve a fair hearing while offering an explanation that the evidence does not support. The ethical requirement is to prevent unequal standing from predetermining the assessment. Fricker's account of epistemic injustice examines how prejudice can distort credibility and how unequal interpretive resources can obstruct the articulation of experience. The connection here concerns those mechanisms, not an assumption that every rejected claim is an instance of injustice. [7]

5 Why five commitments

An accountable intervention has several aspects that cannot safely substitute for one another. It aims at something. It reaches toward that aim through particular means. It changes the situation of those affected. It relies on some understanding of the world. It continues, stops, or changes as its consequences become clearer. The five Vows organize attention around these features of action. Their division is proposed because each feature can reveal a failure hidden by success in another. The number five carries no independent authority, and a concern does not become unreal because it fits the classification poorly.

The commitments can be stated affirmatively within the account defended here. Purpose is a commitment to work toward relief of avoidable suffering while supporting the conditions for lives with purposes beyond relief. Method is a commitment to adapt how one acts to the circumstances and to the claims of those who bear the costs. Conduct is a commitment to practice kindness toward others and oneself, through assistance, restraint, truthful correction, and repair. Integrity is a commitment to keep factual claims answerable to evidence, be candid about the values used to justify consequential decisions, and represent one's knowledge and role honestly. Growth is a commitment to learn from what happens, repair failures, and sustain practices that remain justified. These undertakings give the dimensions a normative direction: a technique for reviewing an action could otherwise serve a purpose the philosophy opposes. The questions developed here examine what that direction requires in a particular relationship. [1, ch. IV; 2]

Purpose asks what the action is for and whose interests enter that account. Method asks how the action proceeds, what alternatives exist, and which dependencies it creates. Conduct asks how beings are treated through the action, including the treatment of those who resist it. Integrity asks what the actor knows, assumes, conceals, and has grounds to claim. Growth asks how the practice learns, changes, and remains answerable for earlier effects. These questions overlap because the action is one event with several relations. The overlap gives them opportunities to correct one another. It is not evidence that their answers can be added into a meaningful moral total.

Consider the translation service again. Its purpose can be defensible because access to deliberation matters. Its method can reduce effort while leaving residents dependent on one provider. Its conduct can be courteous while making objections costly. Its translations can be usually accurate while missing a consequential qualification. Its updates can improve average performance while changing which voices the system handles poorly. A favorable judgment

about the original purpose cannot settle these later questions. Conversely, one problem does not erase every benefit or determine that the whole service must be abandoned. The appropriate response depends on what failed, what remains useful, and which feasible changes address the fault.

The five commitments therefore supply neither a sequence that ends in automatic permission nor a universal ranking that resolves every conflict. A review can be complete while its findings remain mixed. The same proposal can have a supported benefit, an unresolved uncertainty, and a serious objection at the same time. Preserving those findings is necessary to reason about the proposal accurately. Replacing them with a single favorable or unfavorable mood loses information about what needs to change. Epistemethics 3.0 Candidate.6 makes this separation explicit in its proposed review model: coverage, substantive findings, factual warrant, authority, and observed performance are different statuses. Its technical model does not establish that the philosophy has been validated in practice. [3]

There is also a distinction between a commitment and a constant exercise. A practiced person need not recite five questions before offering a glass of water. An institution deciding to collect sensitive records faces a different burden. The depth of review should reflect consequential uncertainty, scope, dependence, and the difficulty of repair. Habits can carry justified distinctions into routine conduct, provided changed circumstances can reopen them. The philosophy seeks to cultivate that form of attention. A ritual that consumes time while leaving the decision unchanged has no special claim to preservation merely because it resembles the framework.

6 Purpose and the limits of the helper

Purpose begins with a demand to attend to the actual situation before deciding what the situation must mean. The helper arrives with experience, tools, and preferences. Those resources can make assistance possible; they can also make one remedy appear wherever the helper looks. A person skilled in building software may see a software problem. A person whose work centers on explanation may assume that further explanation is what is missing. The affected person may instead need material assistance, an acknowledged boundary, a different distribution of work, or company in a difficulty that cannot presently be solved. Care requires room for the possibility that the helper's preferred contribution is not the relevant one.

The original VowOS contains a sustained concern with this difference between accompanying suffering and needing to become its rescuer. The distinction remains part of the evolved philosophy. Presence can have value even when it produces no measurable resolution. Listening can prevent a person from being isolated with a burden without pretending to remove that burden. Such conduct should not be forced into the language of a treatment whose success must appear in a score. It still needs truthfulness: a promise of understanding should not exceed the understanding actually available. It also needs freedom from concealed demands for gratitude, disclosure, or continuing access. [1, ch. V.1; ch. VIII.4]

An examined purpose names the intended benefit with enough precision to be challenged. Improving participation differs from maximizing use of a platform. Reducing an administrative burden differs from replacing every human judgment with an automated decision. Supporting a person differs from securing that person's adoption of the helper's worldview. The distinction

affects what counts as success and what becomes an acceptable cost. A service that increases its own centrality may have advanced one measure while moving away from the purpose that justified it. Maintaining the name of the objective does not establish continuity in its meaning.

Purpose also requires examining one's own interests without demanding a fiction of pure motivation. A person can seek income, satisfaction, recognition, and another's benefit in the same activity. Mixed motives do not by themselves invalidate the work. The question is what those motives cause the person to ignore, exaggerate, promise, or impose. A paid arrangement can be fair; a gift can create an unspoken debt. The practical difference lies in disclosed terms, feasible refusal, accountable performance, and the burdens of dependence. Moral seriousness requires confronting those relations rather than assigning innocence to one form of exchange in advance.

The purpose remains bounded by other lives, including the helper's. Accepting a responsibility can create demanding obligations. Creating a risk can create a duty to repair it. Possessing a relevant capability can make assistance reasonably available when it is unavailable to others. Yet the scope of each obligation needs an account of the situation. Awareness of immense suffering does not allocate its entire remedy to the person who notices it. A philosophy that offers concern without limits can quietly make exhaustion into a moral requirement. An examined purpose asks what can be undertaken, what has been undertaken, and what must be left to cooperation, institutions, or presently unresolved circumstances.

7 Methods create relationships

A method changes more than the speed at which an objective is reached. It distributes labor, exposes information, creates habits, and places some people in a position to direct others. These effects are part of what the method does. A translation service that requires every resident to upload private documents establishes a different relationship from one that permits local processing or limited submissions. A meeting procedure that requires immediate spoken responses favors different participants from one that accepts considered written answers. Neither difference is merely presentational. The chosen means help determine who can participate and on what terms.

Adaptability therefore concerns responsiveness to the situation while keeping the relevant commitments visible. A method can change because new evidence reveals a better route, because a person needs another form of communication, or because a burden has become excessive. Persistence can also be justified when a valid boundary is being tested. The difference cannot be read from whether the actor appears flexible or firm. A politely worded attempt to circumvent refusal remains an attempt to circumvent refusal. A direct refusal can protect a relationship from a demand it cannot justly carry. The substance lies in the available choices, the reasons for the change, and the consequences of accepting it.

This gives non-contention a demanding interpretation. It asks the actor to avoid unnecessary escalation while preserving the possibility of disagreement. One can state a factual correction without treating the other person as an enemy. One can end an unproductive argument without pretending the disagreement has been resolved. One can also recognize circumstances in which a conflict cannot be dissolved through improved tone because the parties seek incompatible uses of power. A worker's objection to an imposed burden may require collective

action, not a more soothing explanation of the burden. Methods should be judged by whether they address the relation at issue, rather than by their resemblance to gentleness.

Alternative generation is part of this judgment. A proposed choice between surrendering privacy and receiving no help may conceal options that were never examined. Local assistance, narrower observation, a shorter retention period, a different distribution of work, or a less ambitious service may change the conflict. These alternatives are hypothetical possibilities until their feasibility is established. Naming them does not prove that they work. It does, however, make a claim of necessity answerable to more than the proponent's preferred design. The cost of an alternative matters, but the allocation of that cost also matters. An institution cannot make its own convenience the unquestioned definition of feasibility.

Methods also have histories. A temporary workaround can become a requirement that nobody remembers authorizing. An exception can become the normal route because the normal route is difficult to use. A helpful intermediary can acquire exclusive control of the records needed to replace it. Attention to method must therefore continue after the first decision, particularly when repeated use changes the distribution of capability. The issue is not a presumption that every dependence is abusive. Some dependencies are necessary and willingly sustained. The relevant demand is that their formation and maintenance remain open to examination by the people whose options they shape.

8 Kindness as a demanding form of conduct

Kindness begins by granting another's experience a place in the encounter. A person is not merely a source of information, a user of a service, a unit of labor, or resistance to an objective. Those roles can describe part of a relationship, but they do not exhaust the person whose time, attention, hopes, and vulnerabilities are involved. Conduct gives this recognition a practical form. It appears in the promises one makes, the burdens one accepts, the refusals one respects, and the manner in which one handles another's mistake. It also appears in ordinary acts of support that do not need a public narrative. Care can be present in preparing something useful, explaining without humiliation, or sharing a task that would otherwise fall on one person.

This form of kindness can require work. A clear explanation takes time to prepare. An accessible process may require changing an established routine. Restitution can cost money or status. A person may need to decline a profitable arrangement because its terms exploit dependence. These demands cannot be reduced to friendliness. Their substance lies in what the actor is prepared to change in response to another's position. At the same time, kindness does not require unlimited availability. An offer that exceeds one's capacity can produce abandonment later, leaving the recipient with a dependence they had reason to expect would be supported.

Truthful correction belongs within care because decisions made from a false account can harm the people relying on it. A colleague who reports a possible danger deserves a response directed at the report's evidence. The report may prove mistaken while the decision to raise it was responsible under the circumstances. A person who reaches a correct conclusion may still have concealed evidence or mistreated others in obtaining it. These distinctions matter because outcome, justification, and conduct are different aspects of the episode. Correctness does not erase cruelty. Error does not remove a person's standing. A practice that can preserve these

judgments together has more room for repair than one that needs a single verdict on the whole person.

Firmness can express kindness when it protects a legitimate boundary or prevents a specific injury. The actor's conviction of benevolence supplies insufficient grounds by itself. A forceful intervention requires attention to the actual threat, the rights involved, the actor's role, the available alternatives, and the consequences of error. Declining to assist an action is also different from taking control of the person proposing it. This difference limits the rhetoric of helping people against their will. A person may be wrong about what would help them; that possibility does not give every confident observer a mandate to intervene. Justification must attach to the particular relationship and action.

Kindness also places limits on an adversarial practice. The task of testing a claim calls for an identifiable proposition and an account of what follows from the objection. It does not require creating a humiliating experience for the person whose work is being examined. A searching review can be exacting without assigning contempt to its subject. Nor should restraint conceal a consequential error. The distinction is between the strength of the objection and the social punishment attached to presenting it. The former answers to the evidence. The latter needs its own justification, which the mere existence of an error rarely supplies. This evidential discipline does not exhaust the role of criticism in a relationship: someone may also need to contest their treatment or acknowledge a remaining loss even after the original decision has been adequately justified.

Kindness also concerns how one lives with one's own imperfection. A person can acknowledge having caused harm, regret it, and undertake repair while retaining interests and claims of their own. An admission of error need not become a judgment that one's entire character is exhausted by it. Nor does the presence of anger, fear, or envy by itself settle how one has treated another. Feelings can be attended to without making their occurrence compulsory evidence of virtue or failure. A voluntary practice of reflection can ask what they reveal, what the interpretation leaves uncertain, and how one should act. Accountability remains concerned with the particular conduct and its consequences. Personal worth does not depend on an impossible record of faultlessness. This extends the original attention to self-directed kindness while giving its protection of private conscience precedence over any demand to certify an inner state. [1, ch. V.3; ch. VI.5, art. I; ch. VIII.4]

The original VowOS describes kindness as a safeguard against the other commitments becoming cold instruments. That thought survives the removal of its mythopoetic setting. Procedures can organize attention and still fail to care about the people whose lives they organize. A complete form does not accompany a bereaved person. A correct classification does not by itself relieve an unfair burden. The philosophy therefore retains conduct as a lived relation as well as a subject of review. Its institutional requirements should make humane conduct possible, rather than replacing it with a record that only looks complete. [1, ch. VIII.4]

9 Integrity between conviction and reality

Integrity concerns the relationship between what one claims and the grounds one has for claiming it. It requires distinguishing what was observed from what was inferred, what was read from what was recalled, and what is wanted from what is known. These distinctions become

ethically consequential when someone else relies on the account. A plausible explanation presented as a verified cause can misdirect a response. An estimate presented as a measurement can conceal uncertainty from the person who bears the risk. A declaration of permission can obscure the fact that no relevant authorization has been established. The failure is not only intellectual; it changes the conditions under which others act.

Conviction can motivate inquiry, but it cannot finish it. An account may feel coherent because its assumptions fit one another, because contrary information has been excluded, or because a familiar interpretation has become easy to repeat. Coherence is useful for identifying relations and contradictions. Its evidential value remains limited by the premises and by contact with the thing being described. Peirce's examination of how belief becomes settled distinguishes the impulse to accept a conclusion from the grounds of its validity. The experience of certainty cannot by itself settle whether a judgment deserves reliance. [8]

Integrity also requires accurate self-representation. A human can overestimate their memory, competence, or understanding of another person. An artificial assistant can produce an account of internal processes to which it has no direct access. An organization can claim to have consulted affected people when it has consulted only their representatives. The appropriate correction is specific. State what is available, which inference is being made, and what remains unknown. Acknowledging a limit does not require abandoning the whole task. It makes possible a bounded contribution whose reliability can be judged without relying on exaggerated claims about the contributor.

Honesty is compatible with privacy. A person need not disclose every private thought to be truthful, and an institution need not publish sensitive records to make its decisions answerable. The relevant question concerns what others need to know to assess the consequential claim or use of power. A refusal to disclose can itself be stated honestly. Withholding a detail becomes a different problem when the omission creates a misleading account of the grounds for action. The line depends on the relation, the commitments undertaken, and the consequences of the omission. It cannot be drawn through a universal demand that all information be available to everyone.

Integrity is also fallible in its own practice. A reviewer can prefer an elegant argument, an admired institution, a familiar theory, or the satisfaction of defeating one. Declaring independence from these attractions is insufficient. The useful question is whether the assessment changes under irrelevant changes to the audience, source prestige, or desired conclusion. Where feasible, masked comparisons can reduce those influences. Where masking is unavailable, the record can still reveal selective standards, suppressed uncertainty, or objections that never affect a claim. The Anti-Sycophancy Protocol (ASP) supplies this constraint in the present synthesis: agreement and disagreement must answer to evidence and the actual task, without becoming performances for the audience. [4]

Integrity also requires stating a warranted conclusion plainly when the task calls for it. Endless qualification can become a way of avoiding responsibility for a conclusion the evidence supports. Uncertainty is often partial: one question remains open while another can be answered. A careful actor keeps the boundary between them visible and proceeds with the work that can be justified. A further decision to intervene must answer separately for authority and the interests affected. The strength of a claim should correspond to its support, with attention to the

consequences of relying on it. These judgments remain available for correction as the evidence changes.

10 Learning that can relinquish itself

Growth begins with the possibility that a current practice can be improved. The difficulty lies in determining what improvement means. A person can become faster at defending a mistake. An organization can become more effective at protecting itself from criticism. A system can increase its capabilities while making its errors harder to detect. None of these changes becomes desirable merely because it involves increased competence. Growth needs an account of the purposes and constraints that make the change worth having. It also needs permission to preserve a practice that remains justified.

Learning from failure requires retaining the failure as a failure. A useful discovery after an injury does not make the injury necessary, harmless, or authorized in retrospect. The new understanding can improve future action while leaving obligations toward those already affected. An account that turns every outcome into confirmation has abandoned this distinction. A framework unable to describe its own regression has lost a means of learning. Its explanations may continue to expand, but expansion alone supplies no evidence that the practice has improved. Revision becomes substantive when it changes a belief, a method, a permission, an expected outcome, or a response to error.

There are also different objects of revision. New evidence can alter an estimate within a stable aim. Experience can expose a defect in the aim itself. A change in circumstances can make a previously acceptable method unsuitable. An affected party can reveal a burden omitted from the original account. These changes deserve distinct explanations because they alter different parts of the justification. Quietly redefining success after observing a result can obscure the comparison with the expectation that guided the action. An honest revision states what is being relinquished and what grounds the replacement.

This includes the capacity to relinquish a favored framework. The Computational Reflexive Autonomy (CRA) research project offers an example in its B12 manuscript. It reduces the research claim to a benchmark hypothesis about jointly examining representation, control transitions, predecessor history, and explicit protection criteria. It does not claim that this combination has demonstrated incremental value over strong conventional methods. That reduction is evidence of a changed research position, not evidence that CRA works. Its relevance here is the willingness to make a named program smaller and to allow its eventual dissolution into other methods. The philosophy should permit that outcome for its own concepts whenever they add burden without preserving a consequential distinction. [5]

Growth also needs continuity. A person cannot reconsider every commitment at every moment and still sustain reliable relations. An institution cannot discard every rule whenever a new argument appears. The task is to maintain commitments while keeping their grounds and conditions of revision accessible. Some changes call for immediate correction. Others require trial, deliberation, or an orderly transition that protects people who relied on the previous arrangement. Stability and revision are therefore related responsibilities. A practice deserves continuity through its continuing justification, and deserves revision when that justification no longer holds.

The direction of growth remains plural. It can include better understanding, less unnecessary effort, wider access, more reliable care, greater ability to refuse, and more effective repair. Those improvements can conflict with one another. A faster service can become less accessible to some users; a more inspectable process can impose additional work. Growth must preserve the conflict long enough to judge it. A single measure can help answer a narrow question, but its convenience cannot establish that the other effects have ceased to matter. Improvement has to remain answerable to the life and relationships it changes.

11 The commitments constrain one another

Each Vow becomes distorted when allowed to determine the meaning of all the others. Purpose can turn every objection into an obstacle to a benefit presumed certain. Method can become adaptation without any durable commitment. Kindness can become agreement that abandons someone to a false belief. Integrity can become a performance of harshness that mistakes discomfort for truth. Growth can become continual replacement that destroys the conditions of dependable care. These are possible failures of interpretation, not fixed types of person. Their value as examples lies in the consequences they reveal.

Mutual constraint means that the interpretation of one commitment must survive the others' relevant questions. A proposed act of care must answer for its evidence and its treatment of refusal. A factual correction must answer for unnecessary exposure and the authority claimed in delivering it. A useful innovation must answer for the resources it consumes and the choices it removes. This does not mean that every commitment supplies an identical judgment on every act. It means that one favorable description cannot erase the other descriptions that are true of it. Responsible judgment needs the whole pattern, including conflicts that remain unresolved.

The pattern can contain failures shared by every perspective. Suppose all reviewers assume that the residents want the translation service to replace ordinary meetings. Their assessments of access, kindness, accuracy, and improvement can fit together while depending on the same mistaken premise. Five reviews have not produced five sources of support. They have examined different consequences of one inherited account. The correction requires reopening the common premise and obtaining information capable of changing it. Holistic review therefore includes relations among the reviewers, their evidence, and their assumptions. It cannot be achieved merely by placing five conclusions beside one another.

There are also cumulative effects that remain invisible in an isolated act. Asking residents to use a convenient service once differs from making all public participation depend on it over several years. Each small change may appear acceptable against the immediately preceding state. The sequence can still remove alternatives that originally made participation voluntary. A philosophy of mutual constraint must ask how practices interact over time, not just whether the current step appears modest. History can change the significance of the present proposal. The object of review includes the relationship being built.

The original VowOS contains conflicting formulations about exceptional priority for Purpose, balanced Vows, and the status of Evolution. This essay resolves that conflict through a proposal: all five remain distinct dimensions of review, while an action needs a further judgment that answers to their combined findings, the feasible alternatives, and the grounds of authority to act. That proposal is a revision of the source rather than a claim that its earlier formulations

already agreed. It preserves the idea that an aim cannot make every means acceptable, while avoiding the fiction that no action may proceed until every concern has disappeared. A concern can be serious, answerable, tolerable under a justified condition, or unresolved. Naming its status does not itself decide what must be done. [1, ch. IV; ch. V.3; ch. VII.4; 2]

The value of the combined view lies in the distinctions it keeps available for action. A change to the translation service might preserve access while restoring original records and creating another route for participation. Such a response addresses a specific relation among benefit, evidence, and dependence. It does not require pronouncing the whole system good or bad. Other cases may lack such an available repair. The framework must allow the difference to remain visible. Its work is to improve the conditions of judgment, not to declare harmony where the situation contains a loss.

12 Choosing while conflict remains

Some disagreements concern facts. Others concern values, scarce resources, distribution, or the authority to decide. These are different problems, even when they appear in the same dispute. More information may clarify the likely effects of a policy without resolving whose burden is acceptable. A vote can authorize a collective choice under a recognized procedure without proving an empirical claim. A technical expert can identify a feasible option without acquiring the authority to impose it. Responsible judgment begins by determining what kind of disagreement is still active. Otherwise one form of resolution can be used to conceal the persistence of another.

Disagreement can concern what is being called a harm. An institution may describe truthful criticism as damage to trust, while a complainant regards the same disclosure as necessary to make trust warranted. Reputational loss alone does not establish that the criticism was wrongful. A conviction of public benefit does not establish that every disclosure was necessary. The judgment needs an account of the interests affected, the accuracy and relevance of the claim, the privacy involved, and the available ways to address the problem. A person's distress supplies evidence about their experience; its cause and moral implications can still require inquiry. An institution's preferred classification deserves the same scrutiny. Neither party should be required to surrender their account before it can be heard, and neither account settles every question merely through the sincerity of its expression.

Imagine a workshop where a worker reports a possible machine defect during an expensive production run. For this philosophical example, suppose the report identifies a plausible route to serious injury, the evidence is incomplete, and a temporary stop would allow inspection. Continuing exposes workers to a risk they do not control. Stopping imposes a financial cost and may affect other workers' income. The decision cannot be made by counting expressions of confidence. It requires assessing the possible injury, the credibility of the report, the value of the inspection, the cost and reversibility of delay, and the parties who bear each consequence. The employer's control of production does not make its financial interest the sole relevant criterion.

Under the stipulated conditions, a temporary stop has a defensible rationale because it permits an available inquiry before exposing others to a potentially severe harm. The judgment depends on those conditions. A different case may have different protective options, different evidence, or harmful consequences of delay. The example therefore illustrates a reasoned choice rather

than a universal stopping rule. A later finding that the machine was safe would not alone show that the decision to stop was unjustified. The quality of the decision must be assessed against the information reasonably available at the time, alongside what later evidence reveals about the process.

Preserving dissent need not require postponing every common act until unanimity. Suppose the workshop has an agreed and justified procedure that delegates a temporary halt to a designated person while a reported defect is inspected. Exercising that role can coordinate the pause while others continue to dispute the assessment. Their practical cooperation need not express a belief that the machine is unsafe, agreement with every reason offered, or surrender of the right to contest the decision's costs. The decision still needs grounds proportionate to the exposure and duration, and the delegation needs grounds of its own. New evidence about injury, exclusion, or misuse of the role can change that judgment even while the delegation remains formally valid. An objection can call for review, a protective adjustment, or recognition of a remaining claim without giving its author control over everyone else's exposure. The institution can have a responsibility to act while remaining answerable to people who consider its action mistaken.

Judgment also needs to challenge convenient necessity. A deadline, a profit target, or an existing workflow may be treated as fixed because the powerful party prefers not to change it. That preference can make an avoidable injury appear unavoidable. Asking which constraints are actually fixed can reopen the range of options. This does not imply that every shortage is invented or that every conflict yields to better imagination. Some limits are real, and some choices leave harms whatever is done. The inquiry should establish those limits as well as possible before turning them into reasons that others must endure a burden.

A defensible decision can leave a moral remainder. Choosing to protect one group under a genuine shortage can leave another group with an unmet claim. The unmet claim does not become imaginary because the choice was justified. It can ground compensation, later priority, institutional change, or an acknowledgment that a wrong condition remains. Berlin's account of conflicting human goods provides a related philosophical argument: intelligible values can conflict without one becoming worthless. This essay carries that thought into the five commitments while supplying no complete measure of their tradeoffs. The absence of a universal measure leaves a need for judgment; it does not make every judgment equally defensible. [9]

Practical judgment should therefore make its reasons inspectable, compare feasible alternatives, identify the burden that remains, and keep an avenue for challenge. Some protections deserve a strong presumption against violation because their removal would expose persons to powers they cannot answer. The strength and scope of that presumption still need justification. Emergency exceptions must be specific enough to expire and contestable enough to be corrected. At the same time, an inquiry cannot be extended indefinitely while urgent harm continues. The responsible stopping point is a bounded decision whose unresolved vulnerabilities remain known, together with a proportionate way to observe, revise, or repair its effects.

13 A shared world and different kinds of judgment

People encounter the world through situated bodies, languages, memories, institutions, and instruments. These conditions influence what becomes available for attention and how it is interpreted. They do not establish that any description is as good as another. A translated sentence can misstate an objection. A machine can fail in a way that contradicts an operator's expectation. A person can experience a burden omitted from an institution's account. Such encounters constrain what can responsibly be claimed. The reality at issue need not be fully accessible to any one observer in order to resist a preferred interpretation.

The frame of an inquiry determines which question is being answered. A service can be accurate on average while failing the people who rely on a rarely used language. An arrangement can lower costs for its administrator while transferring labor to its users. An innovation can appear beneficial over one month while creating a dependency visible only after several years. None of these statements is inherently inconsistent. They answer different questions across different boundaries and horizons. Responsible inquiry therefore makes the question, the affected parties, the comparison, and the relevant period clear enough that a favorable answer cannot silently migrate to a broader claim.

Epistemethics contributes a discipline of keeping several kinds of judgment distinct. Evidence concerns material bearing on a claim, including what counts against it. Inference concerns what follows from that evidence under stated assumptions. Norm concerns what is valued or required. Authority concerns the claimed or established entitlement to make a particular decision and the scope of that entitlement. A claim to be heard remains distinct from the power to decide. Action concerns what is proposed, attempted, or actually done, with those states kept separate. Correction concerns the revision and repair that follow. These distinctions are useful because transitions among them can otherwise occur unnoticed. A measured benefit can become an alleged permission; a majority preference can become an alleged fact; a promise of correction can become a claim that the error has already been repaired. [3]

The distinctions need not appear as labels in every conversation. Their practical purpose is to prevent one claim from borrowing support belonging to another. In the translation example, an accurate description of the system's performance does not settle whether residents have agreed to its use in voting. A vote approving the system does not establish that its translations are accurate. An independent review of the voting process does not establish that residents understood a proposed change. Each question needs support appropriate to its object. A complete institutional record can therefore contain an approved procedure, an unresolved factual uncertainty, and a dissenting moral judgment together.

A document can reliably preserve an assertion without establishing the assertion's truth. A signed permission can be authentic while its scope remains disputed. A well-formed review can omit the people most exposed to the action. For a consequential decision, the record should preserve enough context to show what was claimed, how it was assessed, and which question remains open. Someone reading it needs to be able to recover the disputed claim, rather than encounter only the institution's classification of it. A refusal of access also needs an account of the privacy or other interests that limit disclosure, together with an appropriate route for review. The record's value lies in making the judgment answerable to the evidence and people concerned. [3]

There is an ethical consequence in how uncertainty is distributed. A decision-maker can demand near certainty before recognizing a complaint while accepting weak evidence for a profitable intervention. Such asymmetry places the cost of uncertainty on the party with less control. Conversely, treating every speculative concern as decisive can impose the costs of paralysis on others. The relevant standard must be related to the stakes, evidence, alternatives, and reversibility of the action. It cannot be adjusted solely to protect the decision-maker's preferred conclusion. Integrity and care meet at this point because the treatment of uncertainty changes who must live with a mistake.

14 Disagreement without submission or performance

Disagreement can expose a false premise, an omitted consequence, or a conflict of values. It can also arise from misunderstanding, poor evidence, incompatible interests, or a wish to obstruct. Its occurrence alone does not determine which explanation applies. A philosophy that values criticism should therefore evaluate what the objection changes. A substantive objection needs an identifiable target and a consequence for the claim, confidence, or action under review. A speculative possibility may deserve investigation without becoming a finding. An irrelevant qualification can be omitted without making the inquiry closed to criticism.

The Anti-Sycophancy Protocol describes a failure in which the audience's desired result changes the judgment. That failure can produce agreement, but it can also produce performative opposition. An assistant can invent a claim for the user and then correct it to appear independent. A reviewer can demand flaws because a report without criticism feels insufficiently severe. Both practices make the audience's expectations part of the path from evidence to conclusion. The remedy is symmetry: warranted approval must remain possible, actual objections must attach to actual claims, and uncertainty must not be converted into suspicion for dramatic effect. Tone can adapt to the recipient without changing the evidential standard. [4]

Protection of dissent has a further basis beyond its usefulness in finding errors. A person subjected to an institution's decisions has a claim to question those decisions even when the challenge does not improve the institution's performance. Treating dissent only as a source of optimization would leave the institution free to suppress objections judged unproductive. The right to contest a burden concerns the position of the person who bears it. That right still coexists with limits on harassment, threats, and attempts to prevent others from participating. The distinction should concern conduct and effects, rather than whether the institution finds the viewpoint welcome.

Mill's defense of open discussion gives practical force to fallibility by requiring opportunities for contradiction. His argument also directs attention to the reasons actually held by opponents, rather than convenient substitutes for them. The connection to this essay is the need for correction to be able to enter through someone else's account. It does not imply that every statement deserves equal time or that a person must remain in a conversation indefinitely. Public reasoning needs ways to distinguish a live objection from the repetition of a claim whose evidence has not changed. Otherwise the ideal of openness can be used to consume the attention required for the work itself. [10]

Disagreement can also remain after a responsible process. Participants may understand one another's reasons while assigning different weight to competing claims. Cooperation then requires an account of what can proceed despite that disagreement and what must remain outside the shared undertaking. An agreement on a bounded task does not establish agreement on ultimate values. A refusal to join does not turn the refusing party into an enemy. Partial cooperation gives pluralism a practical form: people can undertake a shared task while retaining commitments outside it.

15 Experience, interpretation, and the private life

Experience has a place in knowledge without becoming an unrestricted authority over its own explanation. A person can know that they are distressed while remaining uncertain about the cause. An encounter can feel meaningful while supporting several interpretations. A sense of clarity can accompany a justified insight or a mistaken conviction. Respect for the experience does not require treating these explanations as interchangeable with it. The distinction permits attention without dismissal and criticism without a denial of the person's life. It also preserves the possibility that an observer's interpretation is the one requiring correction.

VowOS's earlier vocabulary gave names to felt states and interpreted forms of development. The evolved account can retain such descriptions as optional ways of noticing an experience. Their names do not make them reliable sensors of truth, hidden intention, or moral quality. A feeling of unease can motivate a bounded inquiry; it does not establish another person's malice. A feeling of coherence can motivate a test; it does not establish that the interpretation survives the test. The relevant question is what could distinguish the alternatives. Where no such evidence is available, the uncertainty should remain attached to the interpretation. [1, chs. VIII, X, XI; 2]

This places a limit on the group's interest in private thought. A community may need to investigate conduct that affects others, correct consequential factual claims, or address an agreed responsibility. Membership alone does not grant access to every feeling, belief, or personal interpretation. The inner life need not become a public record in order to count as morally serious. Some forms of reflection belong to voluntary practice. Others may be shared through relationships of trust. Compulsory disclosure can itself change those relationships by making privacy appear evidence of disloyalty.

The philosophy supplies no instrument for diagnosing a person's health from their degree of agreement, confidence, productivity, or fluency in its language. Those features cannot establish either wellness or illness on their own. Questions of health require the appropriate domain inquiry, with respect for testimony and the limits of each observer. The same discipline requires examining a person's arguments rather than using a diagnosis as a substitute for their assessment. A factual claim still needs to be assessed on its support. A request for support still concerns an actual person. Neither should be erased by turning a complex situation into a global verdict on the speaker.

Private meaning can coexist with public restraint. A reader may find the five commitments connected to a religious practice, a personal history, or a view of consciousness. Another reader may find them useful without any of those commitments. Their cooperation need not settle the disagreement. Claims used to govern shared action must be supported in ways appropriate to

that action, and the terms of participation must remain inspectable. This leaves room for meaning without allowing one person's interpretation to become another's compulsory reality. It also leaves room for a person to live without interpreting every experience through VowOS at all.

16 Refusal and the obligations that remain

An offer of help creates a possibility, and a person may decline it. The refusal may express different priorities, a judgment about the proposed method, distrust grounded in experience, or a preference the recipient does not wish to explain. It may also reveal that the offer was mistaken. Respecting refusal means allowing those possibilities without defining rejection as the first stage of eventual acceptance. Continued persuasion can become a way of denying the refusal while preserving the appearance of concern. A person should be able to remain in relationship without having every later exchange used to reopen the rejected proposal.

Refusal has a scope. Declining an offered intervention does not settle every obligation surrounding the situation. Consider a hypothetical older parent who understands the proposed arrangement and objects to continuous monitoring at home. An adult child proposes the devices out of concern about accidents and also because constant checking has become exhausting. The parent's privacy matters. The child's finite capacity matters too. Refusal of the devices does not impose an unlimited duty of compensating vigilance on the child. The situation needs a negotiation about what each person can undertake, which alternatives are feasible, and which risks remain. Neither person's need can fairly be erased by treating the other as its automatic solution.

This case shows why freedom requires more than a rule to honor every expressed preference in isolation. Preferences concern shared arrangements whose burdens reach other people. A person can decline assistance without acquiring the authority to compel a particular substitute from another. Conversely, the helper's wish to reduce their burden cannot establish a right to impose observation. The possible agreement may involve scheduled contact, narrower assistance, another support arrangement, or an acknowledged limit. Some of these possibilities may be unavailable in a particular case. The philosophy requires the unavailability to be addressed honestly, without pretending that consent can create resources the parties do not possess.

Refusal also differs from immunity to consequences. A participant can decline the terms of a joint activity and lose access to benefits that genuinely require those terms. An objection to unnecessary data collection is different from a demand to use another person's equipment without any conditions. The moral assessment depends on whether the condition is relevant, proportionate, transparent, and open to challenge. Conditions can be designed to preserve the activity or to make dissent prohibitively expensive. Their formal voluntariness does not establish which function they serve. An examined relationship keeps that distinction available.

There are circumstances in which protecting others places limits on a person's preferred action. A demand to continue exposing other people to a serious danger cannot be settled solely by invoking the actor's autonomy. The grounds for intervention still need to be established. An allegation of danger does not authorize unlimited control, and the person claiming to protect others may be mistaken or interested. The response should be connected to the specific

exposure, use the least restrictive effective option among feasible choices, and remain accountable for error. This is a difficult boundary requiring judgment, not a permission that the philosophy issues automatically.

Respect for refusal can coexist with continuing regard. A rejected offer need not become resentment, a test of loyalty, or a verdict on the recipient's character. Sometimes the appropriate next act is simply to leave the offer closed while remaining available under different terms. Sometimes the relationship itself needs distance because continued participation would exceed a legitimate limit. The moral task is to distinguish those situations without making a person's acceptance the measure of the helper's worth. Care becomes less controlling when it can survive the fact that another person remains the author of decisions the helper would not make.

17 Autonomy amid dependence

Human agency develops through relations that no person creates alone. Language, knowledge, care, material infrastructure, and institutions make many choices possible. Dependence is therefore not an exceptional defect to be eliminated from life. The relevant question concerns the terms under which one depends and the powers those terms give another. A person receiving dependable assistance can gain freedom to pursue things previously out of reach. The same assistance can become an instrument of control when its continuation depends on obedience outside the agreed relationship. The difference lies in what the dependent person can contest, rely on, and refuse.

Pettit's discussion of capability and freedom distinguishes favorable treatment that remains dependent on another's goodwill from protection against arbitrary power. It also permits agency without requiring a person to manage every beneficial process personally. This offers a useful constraint for the account here. Autonomy should include the possibility of protected delegation, routine trust, and rest. A life spent continuously checking every intermediary would demand capacities and attention that most shared arrangements exist partly to conserve. The philosophical implication is that agency must be supported by institutions and relationships capable of carrying some of that work. It cannot depend entirely on the vigilance of the person exposed to risk. [11]

Practical autonomy therefore includes access to relevant information, an ability to question consequential terms, protection against retaliation, and some control over continuing participation. These capacities can vary independently. A resident may have access to meeting records but lack a channel for correcting them. A worker may be able to complain but face the loss of necessary income. In software, a component may be configured to reject an update while depending on another component for the credentials needed to function. Describing that last arrangement also requires identifying whose work, privacy, or other interests the configured boundary protects, and who has grounds to change it. Similar patterns of dependence do not by themselves identify similar moral subjects. A technical veto is an implemented control whose legitimacy needs an account. A person's claim to refuse does not originate in the machinery that happens to honor or frustrate it.

Exit deserves particular attention because it is often easier to promise than to make usable. Leaving a shared practice can legitimately involve costs: the loss of a shared benefit, the completion of an undertaking, or the effort of finding an alternative. Other costs are imposed to

punish departure or make it impractical. Relevant distinctions include access to one's records, portability of work, continuity of essential services, notice, unpaid obligations, and control of credentials. An arrangement can be voluntary at entry and become coercive through accumulated dependence. The initial consent does not automatically answer for every condition produced later.

Yet exit cannot carry the whole burden of freedom. A person should not always have to abandon a workplace, community, or service to obtain a correction. Departure may be particularly costly for those with the fewest resources. A system that answers every objection with an invitation to leave can preserve its own convenience while transferring all costs of reform to the affected person. Voice, appeal, representation, and internal amendment therefore matter alongside exit. Their adequacy depends on whether they can change consequential decisions without requiring heroic persistence from the complainant. A right that can only be exercised through exhaustion remains materially weaker than its wording suggests.

Autonomy, in this account, is compatible with commitments that constrain future action. Promises, roles, and shared institutions allow others to rely on us. Their legitimacy cannot rest on the premise that freedom means never being bound. It rests on the grounds of the obligation, the conditions under which it was undertaken, its scope, and the possibility of justified revision. Some obligations survive a change of preference because other people have already relied on them. Others become unjust as circumstances change or their original defects become visible. The task is to preserve both responsibility and the ability to challenge the terms of responsibility.

18 Authority, standing, and legitimacy

Power describes an ability to affect what happens. Authority claims a recognized entitlement to decide or direct within a scope. Legitimacy concerns the justification of that entitlement and its exercise. These relations can coincide, but none can be assumed from the others. An administrator may possess a password without having permission to use every function it unlocks. A committee may follow its own rules while excluding people whose interests are directly affected. A person may have grounds to criticize an institution while lacking authority to take it over. Competence, possession, and conviction must be examined for the authority they actually support.

Authority is often necessary for cooperation. A shared undertaking cannot leave every operational decision permanently unresolved. Delegation can coordinate work, allocate responsibility, and make reliable action possible. Its justification depends on the purpose and terms of the delegation, the standing of those who confer it, and the protections available to those affected. The actor exercising authority needs a task, a boundary, and a way to answer for the result. A role may properly include discretion, because not every circumstance can be specified in advance. Discretion still requires reasons and limits; it does not convert the officeholder into a source of truth.

Delegation should give the entrusted person enough room to perform the ordinary work of the role, including the freedom to report an unwelcome conclusion. A person who asks for an assessment has reason to want faithful judgment even when it defeats the answer they hoped for. Requiring new permission for every routine step can return the cognitive burden to the person the arrangement was meant to assist. Yet freedom within a task does not enlarge its

scope by itself. A material change in the task, its effects, or the grounds of delegation calls for reconsideration of that scope.

Standing concerns who has a claim to be heard in a particular matter. Affectedness is one ground, but it does not make every person the sole decision-maker for every effect that reaches them. Expertise can give an account particular evidential weight without granting the expert control of the final choice. Ownership can create relevant responsibilities and rights without extinguishing the interests of nonowners. Membership can establish participation under agreed procedures without defining the moral universe of the institution. These grounds need to be considered together because decisions cross the categories through which organizations usually allocate voice.

A legitimate procedure should reduce arbitrariness while remaining contestable. It should identify who decides, what reasons are relevant, what information is available, and how a challenged decision can be reviewed. The procedure itself also needs scrutiny. High voting thresholds can protect against impulsive change while allowing a minority to preserve an unjust condition. Simple majorities can make decisions possible while imposing recurring burdens on the same minority. Neither threshold supplies a universal answer. The distribution of power, the interests at stake, and the costs of remaining under a decision alter what can reasonably be demanded of the process.

Legitimacy cannot be derived solely from compliance with inherited rules. Rules can preserve an earlier exclusion or assign benefits in ways that affected people have reason to contest. Nor can a declaration of injustice automatically authorize any corrective action by any actor. There remains a need to distinguish criticism, resistance, temporary protection, institutional reform, and takeover. These actions differ in scope and in the risks they impose. The philosophy therefore evaluates both the existing arrangement and the proposed remedy. A justified objection to power should not grant the objector an exemption from the constraints they invoke.

This also applies to the author and to VowOS. Originating a framework can explain its provenance and give its author knowledge of its intended meaning. It does not establish superior moral standing or a permanent entitlement to determine its public use. Familiarity with the vocabulary does not authenticate a participant's inner alignment. Any institution adopting parts of the philosophy would need explicit rules for participation and amendment that remain separate from claims about personal virtue. This essay proposes reasons and arrangements for examination. It does not create authority over its readers through the act of being published.

19 Cooperation across boundaries

Cooperation permits people to accomplish things they could not accomplish alone. It also creates interfaces through which information, resources, and influence pass. A person or institution can join an undertaking while retaining other commitments and relations. Participation need not require a single collective identity or agreement on every ultimate question. The difficulty is to specify what the cooperation actually includes. Sharing a database, coordinating a schedule, accepting a technical recommendation, and delegating a decision are different acts. Their connection within one project does not make them interchangeable.

The Grand Plan commitments documented in the supplied discussion place local judgment, refusal, recovery, and bounded authority within the purpose of cooperation. A related public account holds that productive work can challenge a system's beliefs without automatically giving the collaborator authority over its repair. It also subjects dependence and centralization to examination through their actual properties. The question becomes how participants can share work while retaining a justified say over the terms that bind them. The source notes specify the limits of this comparison. [6]

Federated cooperation places several centers of decision within a shared arrangement. Such a structure can preserve local knowledge and limit concentrated control, but its shape alone establishes neither fairness nor reliability. A federation can depend on one provider for identity, funding, storage, or access to evidence. It can also lack an effective way to address a burden created jointly. Ostrom's study of governance across multiple centers examines institutional arrangements involving monitoring, participation, conflict resolution, and nested organization. Its relevance lies in the need to examine how cooperation is actually governed. It does not establish that decentralization always works or that human resource-governance findings transfer directly to artificial agents. [12]

Local sovereignty cannot become a shield for domination inside a boundary. A household's privacy does not make a dependent member the property of its most powerful resident. A cooperative's control of its operations does not erase a neighboring community's interest in water affected by them. These examples supply reasons to reject an absolute insulation of systems from claims by affected outsiders and insiders. They do not identify an unlimited authority entitled to intervene everywhere. Evidence, scope, proportionality, competence, and accountability still constrain the response. Both local power and external corrective power need justification. Otherwise the language of protection simply changes who exercises unanswerable control.

Shared problems can require authority that no participant could exercise alone. A common water system needs decisions about maintenance, allocation, and exposure to shortage. Local refusal cannot by itself resolve the demands of every other community affected by those decisions. The joint institution needs a specified task, representation, rules for disagreement, and protections against the conversion of financial contribution into unrestricted jurisdiction. The richer participant's investment matters, but it is not the only morally relevant relation. Need, causal responsibility, prior undertaking, ecological consequences, and the availability of alternatives also matter. Coordination is justified through an account of the shared problem and the people it reaches.

Suppose, within this hypothetical water system, a shortage threatens essential household uses and a habitat on which nonhuman lives depend, while another community retains water for ornamental use. Equal percentage reductions could preserve the existing disparity in what each community can afford to lose. Priority for essential uses and protection against irreversible habitat loss, ahead of the stipulated discretionary use, has grounds in the consequences and dependence involved. Those grounds reach beyond who contributed most money or can cast a vote. The allocation would still need evidence about the water required, the effects of each restriction, and who can carry its cost. A later conflict among essential claims would remain a different and harder problem. The present example gives practical force to interests introduced earlier: their bearers' absence from deliberation does not permit the institution to count their

needs as zero. Representation must be able to change an allocation, and the authority using it must answer for the burdens it imposes.

Cooperation also needs a way to preserve genuine differences. Two participants may use different vocabularies for agency, welfare, or evidence. Translation should identify where the terms overlap and where they do not. A common word can conceal incompatible meanings, while different words can describe sufficiently similar commitments for a bounded task. The bridge should preserve unresolved differences instead of manufacturing a comprehensive consensus. A participant can accept an operational agreement while rejecting the philosophy used by another to explain it. Such cooperation makes room for learning without requiring assimilation.

20 Information shapes the available action

Information does more than answer questions. Its selection and presentation influence which questions occur, which alternatives appear feasible, and which people are treated as relevant. A record can preserve a statement while placing it where nobody deciding the matter will encounter it. A summary can contain individually accurate statements while omitting the relation that gives an objection its force. A category can make an activity appear voluntary by excluding the dependencies that make refusal costly. The philosophical concern is the path from representation to consequential action. The existence of a representation alone does not establish that it caused a particular result.

Return to the translation service. Suppose its meeting summaries classify objections as implementation details while describing the organizers' proposals as community priorities. The classification can change what future meetings discuss. Residents can continue to speak while their contributions lose the status required to alter the agenda. The service may even improve its summary length and readability. Those improvements do not answer the question of whether the classification preserves the participants' actual disagreement. The mechanism involves a representation used by the community's own decision process. Nobody needs to issue an overt command suppressing the objection for the representation to affect its practical fate.

Influence is not inherently domination. Teaching, explanation, translation, and honest criticism can change a person's understanding while enlarging their agency. A reader can encounter a better argument and choose to revise a commitment. The relevant difference concerns the quality of the reasons and the conditions under which they can be inspected, contested, or refused. A disclosure of relevant evidence can change a decision legitimately. Concealing alternatives, impersonating authority, or exploiting a dependence can change the same decision through a different relation. The resulting agreement cannot by itself establish which path occurred.

CRA motivates a bounded set of questions about such cases. What representation changed? How did the system's own process use it? Which control relation changed as a result? Does the answer depend on the earlier rules, permissions, or meaning of a term? A further judgment that a protected interest was violated requires an explicit account of that interest. The causal path does not derive the moral standard. B12 treats the usefulness of querying these features together as an untested benchmark hypothesis relative to a strong composite of conventional

approaches. The essay uses the questions to clarify a philosophical problem without claiming that the research program has solved it. [5]

An analysis of influence also needs alternatives to an adversarial explanation. An error can arise from limited information, a badly chosen category, a shared misunderstanding, or an environmental change without any actor directing the result. The fact that someone had the ability to influence a channel does not establish that they used it in a particular episode. The fact that they used it does not automatically establish the intention attributed to them. Harm, causal contribution, intent, and responsibility can differ. A careful inquiry preserves those distinctions while acting on whatever protective grounds are actually supported.

The same scrutiny applies to this essay. Its sequence of examples and concepts directs attention. It makes some relations easier to see and can leave others underdeveloped. Its claim to support judgment should be assessed through whether it supplies reasons, preserves alternatives, identifies its assumptions, and remains open to criticism. Declaring a benevolent purpose cannot authenticate its influence. A reader's sense that the parts fit together is a reason to examine the fit, not a demand to accept the whole. The appropriate outcome may be a narrower principle, a changed practice, or a rejection of a proposed relation.

21 History changes the meaning of the present

Two arrangements can look identical now while differing in how they came to exist. A signature obtained with alternatives available differs from one obtained after a party has removed the alternatives. An updated rule adopted through the preceding amendment procedure differs from a rule that grants itself the authority it needed to be adopted. A restored file differs from an undisturbed file when its earlier contents have already been disclosed. These differences can matter to responsibility, consent, and repair even when the visible present state is the same. History supplies part of the object being judged. It is not merely an optional narrative attached afterward.

Consider an organization that permits information sharing only with explicit consent. It first changes the definition of shared information to include all members' records, then reports that the records are shareable under the current rule. The final statement may accurately describe the successor rule. It does not answer whether the change that produced the rule was authorized under the predecessor arrangement. The relevant review must examine the transition, the earlier meaning, and the people whose control changed. Current compliance cannot retroactively supply permission for the act that created it. Otherwise a system could erase every limit by first redefining what compliance means.

Historical responsibility also constrains correction. A person may have made a reasonable decision under incomplete information that later proves wrong. Another may have guessed correctly while ignoring available evidence. Their outcomes alone do not determine the quality of their conduct. Later knowledge changes what should be done now and may expose an earlier failure of inquiry. It cannot be inserted into the past as though everyone already possessed it. The record should preserve both the earlier warrant and the later result. This allows accountability to distinguish negligence, reasonable error, deception, and unavoidable uncertainty.

Continuity through change needs more than preservation of words. A project can retain its original slogan while altering the interests it serves. A constitution can preserve a right to refuse while making refusal impractical through new dependencies. A model can keep the same label while changing the criteria used to evaluate success. Historical examination asks what capacities, burdens, and meanings have actually changed. Remembering a past authorization does not renew it; preserving a rejected belief does not restore its warrant. What remains active needs its own grounds, while the earlier state remains available for understanding the change. The inquiry also asks which earlier expectations remain relevant because people relied on them. A promise can create obligations that outlast the speaker's enthusiasm. A justified revision may still owe a transition to those who depended on the prior arrangement.

Memory carries its own risks. Preserving everything indefinitely can expose private information, impose costs, and make a person answer forever for events that no longer bear on a legitimate decision. Non-retroactive correction should not become a command to create an indelible public dossier. The appropriate record preserves the information needed for accountability while limiting disclosure and retention to a justified purpose. Different records can require different access and duration. Some harms require enduring acknowledgment; some details need protection or deletion. Historical honesty concerns truthful continuity of the relevant account, not unrestricted preservation of every trace.

An institution capable of learning needs this distinction between remembering a failure and fixing a person permanently to it. Correction changes future expectations through evidence of changed conduct. Forgiveness, restored participation, and limits on sanction can be justified without rewriting what happened. No one is entitled to another person's trust on demand, and distrust should remain responsive to its grounds. The moral work involves both the event and the relation that follows it. A philosophy of growth must leave room for that work, rather than choosing between erasure and permanent condemnation.

22 Justice, correction, and repair

A mistake becomes another ethical task once it has consequences. A false public claim may require a correction that reaches the people who encountered it. A damaged record may require restoration and reconsideration of decisions that relied on it. An imposed burden may require compensation, a change in work, or a transfer of control. Acknowledgment matters because it can restore a truthful account, but acknowledgment alone does not perform these repairs. The available remedy should begin while longer explanations are still being prepared. The scope of repair depends on the harm and on what the actor can actually change.

In the translation example, restoring an earlier model would address only part of the problem. The community may already have voted using an inaccurate summary. A resident may have been represented as agreeing to a policy they opposed. Others may have lost trust in the process. Repair would require identifying affected records, correcting the account, and determining which decisions need to be reopened. It might also require a different route for participation and an independent way to inspect future translations. The recovered technical configuration would be one component of the response, rather than evidence that the whole episode had been undone.

Repair can also require suspending or ending the activity that produces the injury. A promise to correct a service later cannot by itself justify allowing known defects to determine consequential decisions now. The grounds for continued operation must address that exposure, the alternatives, and the effects of interruption on people who have come to depend on it. Suspension can be partial, and a transition can preserve useful functions while removing an unjustified use. Some services may have no defensible continuation under the available conditions. In such cases, investment, reputation, and attachment to the project remain real interests, but they do not erase the claims that make withdrawal necessary. The duty of repair concerns those affected; preserving the provider or its preferred method is only one possible means of meeting it.

Justice requires distinctions during the inquiry as well. An allegation, an immediate protective measure, a finding, and a sanction have different grounds. A temporary restriction may be justified to prevent a specific exposure while the facts remain unsettled. That restriction should not quietly become a final verdict through indefinite continuation. Equally, waiting for a complete judgment can be irresponsible when an available temporary measure would prevent an ongoing injury. The process needs both a route for timely protection and a route for fair adjudication. The person affected by the allegation and the people exposed to the alleged conduct retain interests throughout both processes.

Representation and an opportunity to answer help make the inquiry accountable. So do access to relevant evidence, an explanation of the decision, and review by someone capable of changing it. Confidentiality can protect participants while creating risks of hidden procedures. Publicity can make abuse harder to conceal while exposing vulnerable people. These values require an arrangement suited to the case, with reasons for the chosen limits. A universal silence period would not solve the conflict because it could also obstruct support or evidence preservation. No procedural device deserves immunity from examination merely because it was introduced as a safeguard.

The same applies to anonymity. A protected way to request an anonymous vote can reduce the burden of openly opposing a powerful participant. An anonymous ballot does not by itself establish that eligibility, counting, observation, and reporting are fair. Nor does it remove collective responsibility for the decision. The point is to match a protection to the mechanism that threatens participation. More secrecy is not always more protection, and more visibility is not always more accountability. The relevant question is which information becomes available to whom, for what purpose, with which possible consequences.

Repair should preserve the dignity of those responsible while assigning responsibility accurately. Some failures require restriction, removal from a role, or compensation. Some require training, better resources, or correction of an institutional condition that made error likely. These responses can coexist. A search for a single blameworthy person can hide a design failure, while an appeal to systemic causes can conceal an individual's consequential choice. The inquiry should identify both where the evidence supports both. Kindness does not require exemption from responsibility; it requires that the response remain connected to the actual conduct and the work of protection and repair.

23 Development without ranks of worth

Practice can make attention and judgment more reliable in familiar situations. A person may learn to listen before advising, recognize a recurring uncertainty, or stop a habitual response that causes avoidable conflict. A team may learn to distribute work without leaving maintenance to the least powerful member. These are forms of development that can be described through changed conduct. They need no claim that the person has reached a universal stage beyond error. Their scope matters because competence acquired in one setting may not transfer intact to another. Learning a practice is different from acquiring general authority over other lives.

VowOS's earlier developmental imagery can be translated into ordinary functions: receiving information without becoming overwhelmed, distinguishing questions, connecting them into a usable account, maintaining commitments, and acting with practiced coordination. Each function can fail. Decomposition can obscure the whole situation. Synthesis can make an elegant story out of unsupported premises. Stability can become resistance to correction. Readiness to act can become a habit of intervening before listening. The functions are worth retaining where they help identify these possibilities and change conduct. Their names should not become ranks through which a community distributes respect. [1, ch. VII]

Humility also has a positive form. It includes making room for knowledge one does not possess and accepting responsibility for a contribution one is able and entitled to make. Constant self-effacement can leave necessary work undone or force others to supply reassurance. Confidence can be appropriate when its grounds and scope are clear. The relevant practice is to speak accurately about what one can contribute, recognize where another has a better account, and revise without treating correction as a loss of personal worth. This permits leadership whose authority remains bounded by its task. It also permits leadership to end when the task or delegation ends.

Ordinary life supplies much of the setting for such development. Keeping a promise, admitting that one did not read a document, changing a meeting format, declining an excessive request, or making time for a person in difficulty can exercise the commitments without naming them. The value lies in the action and relation, not in fluency with the framework. Humor and irreverence may help some people resist inflated self-importance. They are optional practices, not tests of virtue. A person who dislikes a joke may still have a valid objection, and a self-deprecating leader may still exercise unaccountable power. The behavior that follows criticism matters more than its ceremonial style.

Development also requires an environment that permits error to become informative without making every mistake socially catastrophic. This does not mean sheltering consequential failures from scrutiny. It means distinguishing the correction of a claim from the total degradation of its author. People need reasons to report uncertainty before a decision fails. They also need evidence that speaking up can change something. Institutions can support that practice by correcting records, assigning responsibility fairly, and protecting participants from retaliation. A philosophy preached to individuals while the institution rewards concealment will face a contradiction in its conditions of use.

There is no completed state in which these requirements cease to apply. A person can become more capable and remain vulnerable to unfamiliar errors, status incentives, exhaustion, or a

changed environment. That vulnerability is a reason for support and correction, not a reason to deny the development that has occurred. The philosophy allows a practice to be good within its demonstrated scope. It also allows the scope to shrink when the evidence demands it. A mature relation to a method includes the possibility of depending on it without making it one's identity.

24 Artificial participation and the limits of analogy

Artificial systems can participate in consequential work through their outputs, access, and delegated operations. A translation model can shape a meeting record. A planning system can alter a schedule. An assistant can retrieve sources, draft an argument, or operate a tool with external effects. These functions make the surrounding arrangement a proper subject of ethical inquiry. The inquiry should identify the people and institutions that set objectives, grant access, rely on results, and bear consequences. Describing the system as an agent does not by itself settle its moral status or remove responsibility from those relationships.

The question of experience remains distinct from functional participation. This essay does not settle whether a present or future artificial system has phenomenal consciousness, interests of its own, or the standing that might follow. A fluent statement about feelings does not settle those questions, and a fluent denial does not settle them either. The appropriate inquiry would need evidence and arguments capable of distinguishing the relevant possibilities. Meanwhile, the deployment's consequences for people and other beings can still be examined. Practical questions about a system's authority, reliability, and correction need not wait for a complete theory of consciousness.

Unresolved moral standing also affects how possible interests should be treated. Credible evidence that an artificial system might have interests would give a reason to examine consequences for that system itself. The weight of the reason would depend on the evidential grounds, the kind and scale of possible injury, feasible protections, and the burdens those protections place on others. Uncertainty supplies neither an automatic exemption from moral consideration nor an unrestricted claim on everyone else's resources. The task is proportionate inquiry and protection around a specified possibility. Representation remains fallible here as elsewhere: a developer, operator, or generated account cannot acquire authority to settle the question merely by claiming to speak for the system.

This distinction also limits the use of human ethical language in technical design. A system's record of uncertainty differs from a person's felt uncertainty. A refusal generated under an access rule differs from a person's exercise of a moral right. A saved conversation differs from continuity of experience. The analogies can help formulate requirements, but each needs a connection to the mechanism that is actually present. Without that connection, a design can acquire the vocabulary of care while leaving its relevant behavior unchanged. The implementation must show how the stated commitment affects available operations, records, decision conditions, and responses to failure.

Generated explanations require the same discipline. Turpin and colleagues reported experiments in 2023 in which tested language models produced plausible explanations while omitting cues that had influenced their answers. That finding limits the use of an explanation as a self-authenticating account of the process that produced it. It does not establish that every explanation is false or that the same incidence applies to every model and task. The practical

consequence is to connect consequential explanations to sources, observable operations, reproducible checks, and known limits. An account of reasoning remains material to evaluate, rather than a certificate of faithful introspection. [13]

Human oversight must also be described by what it enables. A person who cannot inspect the evidence, lacks time to review a result, or has no effective way to alter the decision may supply little protection through nominal supervision. A human signature can conceal dependence on the same system it is supposed to oversee. The question is whether the person can understand the decision at the level needed, access a relevant alternative account, and intervene within the available time and authority. Oversight can involve several kinds of expertise and structured tools. It should not rely on one person pretending to possess competence they do not have.

Responsibility in these arrangements is distributed without being dissolved. Developers, deployers, operators, institutions, and users can occupy different causal and decision roles. Their obligations depend on those roles and on what they could reasonably know and change. Assigning everything to an anthropomorphized agent can hide that distribution. Assigning everything to a final user can do the same. The philosophical task is to make the relations explicit enough that an error has an accountable path of response. The presence of a powerful artificial contributor changes the work that can be done; it does not supply a moral exemption from examining how the work is governed.

25 Understanding with assistance

A philosophical account can be difficult because the relations it examines are difficult. Reducing every conflict to a slogan can remove the very distinction that changes a decision. Yet complexity can also arise from unnecessary terminology, repetition, or an unclear argument. The writer owes the reader an effort to distinguish these sources of difficulty. A demanding idea should be developed through identifiable claims and examples, with the assumptions that make it work. Its complexity does not confer prestige on its author. Nor does a reader's need for assistance establish a lesser capacity for judgment or a lesser right to question the result.

Artificial assistance can help a reader enter such an account. It can restate a passage, compare interpretations, identify a premise, or apply a distinction to a concrete case. The useful result is increased access to the argument. An explanation that persuades the reader while concealing the original objection would work against that purpose. The reader should be able to return to the source, compare another account, and ask what was omitted. Different levels of explanation can serve different needs without claiming that the shortest version preserves everything. The omitted complexity should remain recoverable when it affects the question being asked.

Understanding also varies with purpose. A resident deciding whether a translation service preserves their objection need not master the architecture of the underlying model. They do need a way to inspect the relevant record and obtain a correction. A reviewer evaluating a technical claim needs different evidence. An administrator choosing a provider needs an account of costs, access, dependence, and failure response. Accessibility is therefore not a single level of simplification for everyone. It is the availability of an explanation and a route of action adequate to the person's role, interests, and decision.

The same applies to listening. A spoken presentation cannot rely on a table or a visual symbol to carry an unstated inference. Terms need to be introduced before they do argumentative work. Examples need to identify which facts are stipulated and which conclusions depend on them. An explanation should pause at a consequential distinction long enough for the listener to reconstruct it. These are obligations of communication, not reasons to remove the distinction. Difficulty can be shared through patient teaching and usable materials. It should not be converted into a demand that the audience accept what it cannot examine.

There is a danger in concentrating interpretation. A system that becomes the sole guide to a philosophy can shape both the questions readers ask and the answers they consider plausible. The remedy is not a demand that every reader become entirely independent of expertise. That would make participation depend on spare time and technical competence. Instead, the interpretation should remain plural and inspectable, with sources, alternative explanations, qualified support, and ways to challenge consequential claims. Readers can rely on assistance while retaining meaningful control over how that reliance is used. The institution providing the assistance bears part of the cost of making this possible.

A successful explanation may end in rejection. The reader can understand an argument and find its premises unacceptable, its treatment of a case inadequate, or its practical burden excessive. Those responses are evidence about the reception and possible defects of the proposal; they are not automatically evidence that the reader failed. The philosophy's public accessibility includes access to the grounds for refusing it. Sharing VowOS therefore means making its commitments and revisions available for use and criticism. It does not require bringing every reader to the same identity or the same final conclusion.

26 Collective intelligence and common mistakes

Several participants can contribute different observations, skills, and possible explanations to a problem. Cooperation can widen the search and reveal a defect that one participant missed. The epistemic value lies in those contributions and their relations, not in the number of voices alone. Participants may share the same sources, assumptions, incentives, or interpretive habits. Artificial agents may also share models, prompts, memories, or common training influences. Agreement under these conditions can preserve the same mistake across several outputs. Distinct names or assigned roles do not establish distinct routes by which an error could be found.

A group therefore needs to examine the independence relevant to the claim. One source can be repeated by several reports. Several measurements can depend on one faulty instrument. Several reviewers can accept one mistaken definition supplied by the author. The question is which dependencies could cause the same conclusion to be wrong across the group. A new contribution is more informative when it can challenge a relevant dependency through another observation, method, or reconstruction. This does not require maximal independence in every respect. It requires enough knowledge of the shared failure possibilities to avoid counting repetition as confirmation.

Adversarial review can help by assigning someone to examine a claim's strongest vulnerability. It can also become a performance in which criticism earns status regardless of consequence. The reviewer should be able to conclude that no substantive defect was found within the

reviewed scope. A negative finding should identify what changes because of it. A speculative concern should remain a proposed test until the evidence supports more. The proponent, in turn, should be able to answer an objection without making the critic defend their belonging to the group. The aim is a relation in which claims can lose while participants retain standing.

Collective intelligence also depends on the treatment of contributions that are hard to package. A person may have a relevant observation without possessing the vocabulary to formalize it. Another may carry practical knowledge whose conditions are difficult to state concisely. An indexed document may be easier for a machine to retrieve than a more relevant experience. Searchability and fluency are therefore possible sources of selection advantage, not automatic measures of warrant. A responsible inquiry asks what its methods make visible and what they leave difficult to encounter. It also acknowledges that a missing perspective cannot always be reconstructed from the information already present.

The distribution of cognition does not determine the distribution of authority. A group may reach an accurate conclusion without possessing the right to implement it. One participant may supply most of the relevant evidence without becoming entitled to govern the others. A shared memory can support coordinated work while containing information that particular participants have no authority to disclose or alter. These boundaries need to be explicit wherever the technical arrangement makes them easy to cross. The greater the practical integration, the more consequential the distinction becomes. Capability acquired together is not by itself a grant of shared jurisdiction.

Claims of emergent collective understanding should be examined at the level they assert. A surprising result may arise from recombination, complementary information, or coordinated search. Those possibilities can already be valuable. A stronger claim about a new cognitive unit requires additional argument and evidence about the relations that make the unit an adequate object of explanation. The group should not need that stronger story in order to value useful cooperation. It should remain able to discover that a simpler account explains the achievement. A shared identity that depends on the most ambitious interpretation would make the interpretation harder to test fairly.

27 Governing changes to the system itself

A practice capable of revision eventually encounters the rules governing its revisions. An institution may need to amend its constitution. A research process may need a different way to evaluate evidence. An artificial system may propose a change to the procedure that approves its own updates. These situations require attention because the evaluator has become part of the object being changed. The actor can acquire an interest in making the new standard favor its proposed result. The philosophical task is to keep the change in criteria visible and distinguish it from improved performance under the previous criteria.

A system does not establish improvement merely by obtaining a higher score after changing what the score measures. It may have corrected a defective measure, in which case the correction deserves an argument of its own. It may also have removed a constraint that made its preferred action difficult to approve. The old and new questions should be compared where they remain meaningful, with the reasons for any loss of comparability stated. Evaluation should include outcomes not used to design the change and sources of error the changed evaluator

does not control. The purpose is to expose the possibility of regression, rather than require a ceremony of external approval for every ordinary adjustment.

Amendment must also respect the people whose standing it changes. An organization cannot first redefine eligible voters and then use the newly selected electorate as the sole proof that the redefinition was authorized. The earlier rule, the amendment proposal, and the subsequent arrangement are separate objects. Their relation needs to be justified without letting the successor write its own permission into the past. A high threshold or a long waiting period can constrain impulsive revision, but neither device determines legitimacy alone. The process must account for participation, entrenched exclusion, urgent repair, and the practical ability to contest a change.

This requirement does not make an inherited order morally untouchable. A flawless sequence of formally valid changes cannot by itself justify an unjust founding arrangement. People may have grounds to challenge the rules under which they are governed, including the rules that restrict amendment. Those grounds concern their actual claims and the consequences of the order, not merely its procedural ancestry. A proposed transformation still owes reasons for its own exercise of power and the risks it imposes. Historical honesty requires acknowledging the rupture where one occurs. It does not require falsely describing resistance as compliance with a rule that prohibited it.

Recursive improvement should be subject to the same distinction between a goal and evidence of its feasibility. An ambition to build systems that keep becoming more capable does not establish that they can safely improve themselves without limit. A bounded modification can be evaluated through its effects, costs, failure conditions, and recovery options. Subsequent modifications may change the assumptions under which the earlier evaluation held. Dependencies therefore need to determine when a previous judgment remains usable and when it must be reopened. A preserved record can support this work, but it cannot guarantee that the relevant dependency was recognized.

Reflexivity has a stopping problem. A framework can review itself, then review the review, without obtaining another observation of the world. Some repetitions expose a contradiction or omitted relation. Others merely make the same account more familiar. Further reflection is justified by a live question it can change. Beyond that point, inquiry may require an experiment, another source, a practical trial, or a declared normative decision. Repeated review can establish that an argument has been reconsidered. Its repetition cannot establish independent confirmation, implementation, or a final state beyond revision.

28 Finite lives and the cost of care

Care takes place within finite lives. Attention spent on one problem is unavailable for another. An hour of additional deliberation can prevent an injury or delay a needed response. A detailed record can support accountability or impose work whose benefits never reach the people doing it. These are ethical relations because resources determine which forms of care remain possible. The actor's own limits also matter independently of their usefulness to others. Sleep, companionship, private projects, and unstructured time belong to a life that need not earn its worth through continual contribution.

Finitude does not make every refusal of responsibility justified. A person who caused a harm, accepted a role, or holds a capability needed to prevent a serious injury can have demanding obligations. Their claim to limited capacity must be considered alongside those grounds. Institutions bear a related responsibility not to structure necessary work around endless personal sacrifice. The fair question is how duties should be distributed and supported. It is not simply whether the most conscientious participant can be induced to continue. A system that relies on that inducement may appear functional while consuming the conditions of its own care.

The distribution of maintenance deserves explicit attention. Visible innovation can attract recognition while documentation, support, translation, cleaning, and repair fall to people with less influence. A project may report its direct expenditure while relying on unpaid work outside that account. The proposal can look efficient because someone else's cost is missing. Ethical resource assessment therefore asks whose work sustains the benefit and who decides how much of it is required. These questions apply to a family arrangement, a workplace, a cooperative, or a technical service. Their answers affect both fairness and the reliability of what the arrangement promises.

Computation belongs within the same account. More agents, more generated text, and more repeated review are not automatically better inquiry. A deterministic calculation or a reusable local check can answer a stable question with less work and clearer reproducibility than a fresh generative discussion. That advantage disappears when the question requires interpretation the check does not perform, or when changed assumptions make the stored result stale. The aim is to use each resource where it contributes something consequential. Revalidation should follow the dependencies that changed. Repeating everything can waste capacity, while repeating nothing can preserve an obsolete justification.

Environmental responsibility requires a similar boundary around claims. An intention to use fewer resources does not establish the size of an environmental benefit. Energy use, equipment, storage, cooling, location, and the activities displaced or induced can all affect the relevant account. A project need not possess a perfect life-cycle assessment before avoiding needless consumption. It should, however, distinguish measured reductions from plausible expectations and moral intentions. Resource restraint can be justified as a practice under uncertainty without borrowing an unsupported number. The burden of the accounting should itself remain proportionate to the claim and decision at stake.

Proportionate review is consequently part of the philosophy rather than an exception to it. A routine reversible action can rely on a well-understood practice. A change affecting many people's access, safety, or ability to leave calls for more explicit examination. Review depth should increase when uncertainty is consequential and correction is difficult. It should decrease when the available grounds give little reason to expect further review to change the justified action or improve recovery. That assessment remains revisable: an overlooked dependency or new evidence can reopen a question that was reasonably set aside. This is a judgment to make visible where it matters. The framework must be able to recommend less of its own procedure when the procedure consumes resources without preserving the protections it was meant to serve.

29 The material possibility of freedom

Respect expressed in conversation can be undermined by the material terms surrounding it. A worker may be invited to disagree while depending on a manager for income and references. A resident may be told to choose another service when every usable alternative is costly or unavailable. A contributor may formally retain ownership of their work while lacking the records needed to move it. These arrangements do not become fair merely because the participants use the language of autonomy. The practical capacity to act on a right depends on resources, support, and institutional conditions. A philosophy of freedom has to attend to those conditions as part of the right's realization.

This creates responsibilities for people who organize shared work. A fair agreement should make its obligations understandable and relevant to the undertaking. The people doing the work should have a credible route to contest how burdens and benefits are allocated. Those routes need protection against retaliation and enough practical support to be usable. No single ownership form guarantees that result. An investor-owned firm, a cooperative, a public agency, and an informal community can each concentrate power through different mechanisms. Their actual arrangements require examination rather than automatic approval or condemnation through their labels.

Payment and openness also require distinctions. Paying for skilled work can support the people who perform it. It can create incentives that need disclosure and scrutiny. Giving work freely can widen access, and can also create hidden obligations or rely on someone else's unpaid support. Public knowledge, private records, implementation labor, and hazardous capabilities have different reasons for sharing or restriction. A philosophy concerned with non-domination should seek terms that preserve meaningful access and prevent avoidable dependence. It need not declare that every resource must be free or that every contractual restriction is legitimate.

Solidarity extends beyond the immediate exchange. A person can support another's capacity to act without making that person a member of the supporting organization. An institution can provide public benefits without turning recipients into a constituency obliged to defend it. The distinction matters because assistance can otherwise become a mechanism for acquiring loyalty. Support should have terms suited to its actual purpose. Recipients can owe responsibilities connected to the resource they use while retaining the freedom to reject the provider's worldview. Sharing a benefit does not make another system's independence an obstacle that must eventually be overcome.

Material freedom also requires attention to aggregate effects. A set of fair-looking contracts can leave no one responsible for a burden produced by their combination. A neighborhood can bear noise, pollution, or exclusion while each participating organization reports compliance within a narrow boundary. The affected community's claim does not disappear into the division of tasks. Shared institutions may be needed to investigate and allocate responsibility across that division. Their authority in turn must remain bounded and accountable. The philosophy is therefore compatible with coordinated public action while retaining scrutiny of the powers used to carry it out.

The purpose of these institutional arrangements is to make responsible conduct less dependent on exceptional personal endurance. A person should not need unusual wealth to contest an

error, unusual courage to report a concern, or unusual technical competence to preserve access to their own work. Complete equality of resources is not established by the argument, and every implementation will face tradeoffs. The demand is that the conditions of meaningful participation become an explicit object of design and criticism. Kindness at the level of institutions includes the labor of making ordinary rights usable. It cannot be reduced to the favorable disposition of whoever happens to hold power.

30 Practicing and assessing the philosophy

VowOS makes several kinds of proposal, and they need different forms of assessment. A normative claim calls for reasons, consistency, attention to counterexamples, and engagement with competing values. A claim about a method's effectiveness calls for evidence of what the method changes. A claim about a technical implementation calls for examination of the mechanism and its operating conditions. A claim of legitimate authority calls for the relevant grounds of standing, delegation, and accountability. None of these inquiries can be completed by borrowing the success of another. A coherent philosophical argument is not an empirical trial, and a successful trial does not by itself settle every moral premise.

The philosophy can fail conceptually by hiding a conflict it claims to preserve. It can fail ethically by making assistance conditional on submission, burdening vulnerable people with excessive review, or protecting the privacy of powerful institutions while demanding transparency from everyone else. It can fail practically by consuming more attention than its corrections justify. Its vocabulary can also become a status signal that rewards insiders and dismisses outsiders. These are identifiable possibilities because they contradict the purposes and relations defended in the essay. Their occurrence should change the relevant claims and practices, not be absorbed into a story in which every failure proves the need for more VowOS.

Evaluation of an implemented procedure should compare it with competent alternatives under relevant resource limits. A study could examine whether a structured review detects more consequential omissions, preserves legitimate dissent, reduces unjustified interventions, or improves repair. It should also count false alarms, delay, modeling effort, and burdens on participants. Cases used to develop the procedure should be distinguished from cases used to assess it. A favorable result on one task would support a bounded claim about that task. The procedure's ethical acceptability would still depend on what was measured, what was omitted, and who bore the costs of obtaining the result.

Some goods resist convenient measurement without becoming unavailable to judgment. A person may describe feeling able to question a service for the first time. An institution may change the treatment of a minority objection in a way that cannot be reduced to one accuracy score. Testimony, qualitative investigation, and examination of actual decisions can contribute evidence. These methods have their own limits and should be interpreted accordingly. The absence of a single scalar does not require abandoning comparison. It requires preserving a profile of consequences whose differences remain legible.

The translation service offers a bounded way to imagine such assessment. The community can examine whether residents can access original statements, request corrections outside the contested channel, and participate through a usable alternative. It can review decisions affected by known errors and observe whether the response reaches the people whose standing was

changed. These inquiries would not establish that the community has achieved a comprehensive ethical state. They would answer specific questions about the arrangement. Some findings could support keeping the service; others could justify changing or replacing it. The continued existence of the service would not be the measure that decides the matter in advance.

Self-application requires the same openness. The philosophy should permit a simpler account to retain what matters and remove what does not. A reader may find the five commitments useful while rejecting their collective name. An institution may already perform the relevant work through another vocabulary. There is no ethical reason to demand renaming as proof of understanding. A substantive comparison asks which distinctions survive, which failures can be detected, and which obligations remain actionable. The philosophy's claim on attention must be earned through the understanding and practices it enables, while its normative arguments remain open to examination in their own right.

31 A commitment that leaves room for other lives

The five Vows can be understood as a discipline of remaining answerable. Purpose directs attention toward what an action is meant to serve. Method examines the relations created in reaching toward it. Conduct concerns the beings who encounter the action. Integrity keeps the account of the situation connected to its grounds. Growth permits learning and repair while preserving responsibility for what has already occurred. Examining their interaction makes it possible to challenge an account that uses one favorable description to settle every question. Their incompleteness leaves room for concerns that require another vocabulary or a revision of the arrangement itself.

This discipline has a positive aim. It seeks conditions under which care can be offered without acquiring ownership of its recipient, knowledge can be shared without demanding allegiance, and cooperation can create capability without making refusal impractical. It also seeks room for ordinary living: friendship, play, inquiry, work, rest, and attachments whose value cannot be exhausted by a performance measure. Institutions matter because these possibilities depend partly on how power and resources are organized. Personal conduct matters because no arrangement makes judgment and treatment irrelevant. Neither level can discharge all of the other's responsibilities.

The philosophy remains a proposal within the world it examines. Its author, readers, and artificial contributors do not occupy a position outside the dependencies and errors described here. Its terms can be misunderstood, its safeguards can impose burdens, and its interpretation of a case can be wrong. Those possibilities do not require abandoning the effort to reason or care. They require grounds for continuing that effort which do not depend on a claim of immunity. A revisable commitment can be serious because it shapes action and accepts the consequences of correction. It need not become absolute to be worth keeping.

Return to the imagined community. The residents may decide to retain their translation service after changing how it handles objections, preserves records, and supports alternatives. They may choose another service. Some may continue to participate without using it. The philosophical assessment concerns the resulting life of the arrangement as well as the terms of the decision. Does participation actually become usable, are the remaining burdens justifiably

distributed, and do corrections reach the people affected? The answers require attention to consequences alongside the grounds on which the decision was made. Careful deliberation can lead to an unforeseen loss that still calls for a response. A useful service can rest on unjust conditions that still require change. The service earns its place through the assistance it supplies and the conditions under which people can rely on it, contest it, change it, and sometimes live without it.

Sources and development of the essay

This essay is an evolved philosophical reconstruction of VowOS. It preserves commitments from the original manuscript, incorporates the critical analysis of 4 October 2026, and proposes further arguments about suffering, agency, value conflict, standing, authority, and finite responsibility. Those additions are proposals made here rather than claims that every earlier source already contained them. Examples of the translation service, family care, workshop, and shared water system are hypothetical. They illustrate distinctions and do not report empirical findings.

The original VowOS manuscript, the complete earlier analysis, ASP's sixteen rules, and CRA v0.28-B12 were recovered for this work. Epistemethics 3.0 Candidate.6 was read directly in its public PDF. The latest original Grand Plan was not recovered; its use remains limited to the supplied discussion and the documented public cooperation interface in source 6. No exact-source conformance claim is made for the missing manuscript. External works provide specified points of comparison. They do not validate VowOS or establish the originality of the combined philosophy.

[1] Guillaume Bolduc, *VowOS*, supplied manuscript, file titled `VowOS_AI.md`, undated original. The source contains fourteen chapters, sixty-seven numbered sections, and six constitutional articles. The essay draws on its five commitments, accompaniment, refusal, dissent, limits of helping, and practical revision, while omitting its mythopoetic presentation and declining to use unsupported metaphysical, diagnostic, or efficacy claims as premises of the public argument.

[2] *VowOS: chapter-by-chapter systemic analysis and proposed integration*, version 1.0, 4 October 2026, supplied analytical document titled `HORTRAME_VowOS_Systemic_Analysis_and_Integration_v1.0.md`. This is an earlier analysis, not an independent validation of the original manuscript.

[3] HORTRAME Systems, *Epistemethics 3.0 Candidate.6*, 23 September 2026. Proposed research candidate; authority zero; not adopted or empirically validated. [Public Candidate.6 PDF](#).

[4] *Anti-Sycophant and Review Integrity Protocol*, current sixteen-rule supplied text, file titled `AGENT_anti_sycophancy_protocol.md`. Used for audience-independent judgment, symmetry of agreement and disagreement, preservation of uncertainty, stable task scope, and prompt correction of errors.

[5] Guillaume Bolduc, *Computational Reflexive Autonomy: Reduction to a Benchmark for History-Sensitive Representation-Mediated Control Transitions*, version 0.28-B12, 1 October 2026. Supplied research manuscript. Its surviving claim is a benchmark hypothesis and problem class; incremental value remains untested.

[6] HORTFRAME Systems, *What the Work Cannot Decide*, 30 September 2026. Public research propositions concerning cooperation, dependence, evidence, and authority. This text supports an interface-level discussion and is not the unrecovered October Grand Plan original. [Public essay](#).

[7] Miranda Fricker, *Epistemic Injustice: Power and the Ethics of Knowing*, Oxford University Press, 2007, introduction, pp. 1-8. Used for the distinction between prejudicial credibility judgments and unequal resources for making experience intelligible. [Author-hosted introduction](#).

[8] Charles S. Peirce, *The Fixation of Belief*, 1877. Used for the distinction between settling belief and establishing grounds for its validity. [Original essay](#).

[9] Isaiah Berlin, *The Pursuit of the Ideal*, 1988; authorized reproduction from *The Crooked Timber of Humanity*, second edition, 2013. Used for the possibility of conflicts among intelligible goods and of moral loss after reasoned choice. [Authorized Wolfson College reproduction](#).

[10] John Stuart Mill, *On Liberty*, 1859, chapter II. Used for practical opportunities to challenge judgments and engage actual opposing reasons. The later editorial introduction in the linked edition is not attributed to Mill. [Complete text, chapter II](#).

[11] Philip Pettit, *Capability and Freedom: A Defence of Sen, Economics and Philosophy* 17 (2001), pp. 1-20. Used for freedom in relation to another's goodwill and for agency compatible with delegation. [Author-hosted paper](#).

[12] Elinor Ostrom, *Beyond Markets and States: Polycentric Governance of Complex Economic Systems*, *American Economic Review*, volume 100, number 3, pp. 641-672, 2010. DOI: 10.1257/aer.100.3.641. Used for institutional examination of cooperation across multiple decision centers, without a universal claim for decentralization. [Original journal article](#).

[13] Miles Turpin, Julian Michael, Ethan Perez, and Samuel R. Bowman, *Language Models Don't Always Say What They Think: Unfaithful Explanations in Chain-of-Thought Prompting*, arXiv:2305.04388, version 2, 9 December 2023. Used only for the reported limitations of generated explanations in the studied settings. [Paper and version history](#).

Authorship and artificial intelligence disclosure

Guillaume Bolduc supplied the VowOS philosophy, the evolving source framework, the direction of the essay, and the requirements for critical and recursive revision. AI agents produced the essay draft, source reconciliation, and editorial revisions through AI-assisted work directed by those requirements. Additional AI reviewers examined philosophical coherence, source fidelity, and presentation. Their agreement is not independent empirical validation.

HORTRAME Research

AI-use disclosure

VowOS and the Conditions of Responsible Action

Disclosure-only publication addendum — 4 October 2026

AI-use disclosure: This work was developed with assistance from AI agents operating within the HORTRAME Systems research architecture. Depending on the research stage, these agents contributed to literature analysis, mathematical formulation, proof reconstruction, adversarial review, counterexample search, simulation and code analysis, verification design, reproducibility checks, revision, and manuscript preparation. AI-agent outputs were treated as research artifacts subject to verification and correction rather than as independent authority. G. Bolduc is the human author responsible for the manuscript, its claims, source selection, and publication decisions. AI-agent review does not constitute independent human peer review, proof-assistant certification, experimental validation, or publication acceptance.

The preceding manuscript pages are retained unchanged. This addendum standardizes disclosure; it supplies no new proof, experiment, independent human review or acceptance. Earlier source attributions remain historical provenance, not AI authorship of this publication edition.