

Peer Review File

Manuscript Title: People systematically overlook subtractive changes

Editorial Notes: *none*

Reviewer Comments & Author Rebuttals**Reviewer Reports on the Initial Version:**

Referee #1 (Remarks to the Author):

This manuscript examines the possibility that subtractive changes to an existing artifact are less cognitively available and thus less likely to be generated or employed than additive changes. The authors show this effect is general across multiple contexts and cultures. These results are fundamental and are therefore of immediate interest to many people in my own discipline and beyond.

I have one primary concern and one recommendation for this manuscript. The concern is that the prevalence of this effect, though impressive, is contingent on some sort of norm or predisposition. There are certainly situations where subtractive changes are top of mind, so one question is how general the effect really is. Shown a picture of a hoarder's home, a cancerous growth, a flooded field, a sprawling suburb, etc., I believe subtractive improvements would readily come to mind for most of us. It is possible that standard and norms are an important moderator. The authors touch on this in S9 and S10 and line 88-91, but I personally am curious about the limits of the effect because it would help me understand more fully why the effect occurs. Furthermore, I am not sure that many contexts would allow for balance in additive and subtractive thoughts. In a few studies (in the main text, 2-4), it is important to recognize the proximity of the lower limit on subtractive changes relative to that of the upper limit on additive changes. Yes, there are things that could be removed from the golf stimuli, but there are an infinite number of things that could be added.

I personally really like this work, and see it fitting in with work on query theory by Eric Johnson and colleagues. My recommendation is related to that work. I think that the authors should consider an alternative way of analyzing their data in studies 2-4. Given the authors' goal of determining the prominence or availability of subtractive changes, I believe an analysis more similar to that of Johnson, Haubl, and Kienan's (2007) query theory might be more appropriate than the current one in studies 2-4. If I understand correctly, the authors want to illustrate the relative delay in generating subtractive changes, but they don't care about the number of changes overall, for instance. They also are curious if there is a treatment effect on this delay. However, there is additional information and an important distinction between a participant who generates her first subtractive change on her 5th and final try versus the participant who generates his first and only subtractive change on his 5th out of 30 tries. To capture that, the analysis using median ranks utilized by Johnson et al could be helpful. They calculated a standardized median rank difference (SMRD) to capture and test the relative differences in availability of certain thoughts. Applied to the present data, this statistic would be calculated for each participant as $2 * (\text{Median rank of additive changes} - \text{Median rank of subtractive changes}) / \text{number of changes suggested}$, but see the cited paper and particularly the footnote within that paper for important details.

Referee #2 (Remarks to the Author):

The authors propose that, when looking for improvement, people are biased toward additive changes at the expense of subtractive changes - and that this happens in part because subtractive changes come to mind less easily (are chronically less accessible).

I think the authors do a good job of demonstrating that people are indeed biased toward additive changes, but I don't think that they have demonstrated that this is due to subtractive changes being chronically less accessible.

The first set of "motivating observations," which are just intended to demonstrate the phenomenon rather than nail down the process, actually illustrate some of the many reasons why people are biased toward addition. That is, they can all be explained without having to invoke a novel explanation based on differences in chronic accessibility. For instance, the fact that members of a university community propose far more additive than subtractive changes when queried by an incoming president can be explained by (1) political/social reasons: it can be problematic suggesting shutting down an initiative championed by someone else, (2) assumptions that "if it is there, it is there for a reason," (3) sunk cost bias (we already invested so much in these programs), (4) wanting to be recognized for creativity and effort (proposing a new program signals more creativity than proposing shutting one down), etc... The main reason, though, which I think applies to every example, is that people assume they are SUPPOSED to add: the president is asking them to come up with new ideas, not shoot down old ones. In other words, the bias toward additive changes can be accounted for by norms and expectations.

The second set of studies, the actual experiments, aim to go beyond demonstrating the phenomenon (which seems real but also multiply determined) and showing the process (subtractive solutions being less accessible).

Experiments 1-4 do this by demonstrating that people are more likely to offer subtractive changes if they are explicitly told (cued) that this is an option. This doesn't demonstrate the role of chronic accessibility as it is fully consistent with the norms explanation (or, in experimental terms: the cue manipulation activates a demand/ conversational norms effect: it tells participants that subtractive changes are allowed, something many may not have assumed otherwise).

Experiments 5-7 examine the mechanism by showing that subtractive solutions are even less likely when people are under cognitive load. I like this approach more than that of expts 1-4 as it doesn't manipulate norms or expectations (the way the cuing manipulation did). The problem with these studies, however, is that the results are much weaker than for the previous studies. My take-away is that these studies do illustrate that subtractive solutions are less likely to come to mind initially than additive ones, but that this only contributes minimally to the (rather robust) phenomenon that people are more likely to offer additive solutions.

Conceptually, I was surprised that the authors do not offer an explanation for why subtractive solutions are less likely to come to mind. The authors explicitly separate the accessibility (consideration) explanation from the fact that people may PREFER additive changes for a variety of reasons (e.g., see the reasons I specified above for the additive university improvements). However, rather than assuming that subtractive changes are inherently less accessible (for an unknown reason), I think it is more likely that people are simply less likely to first go that route (i.e., consider them) because they usually do not want to. That is, preference drives consideration (as it usually does - it is odd to separate them). In fact, I would argue that subtractive changes are actually easier to imagine than additive ones - as it doesn't require any creative search. This is one of the reasons why people are more sensitive to errors of commission than to errors of omission: it's much easier for us to generate the counterfactual that something had not happened than to generate the counterfactual that something did happen.

I want to be clear: I do think that the phenomenon itself, that people are more likely to look for additive than subtractive improvements is an important and likely robust phenomenon. However,

to examine it, I would suggest avoiding studies that can be explained by demand/norm effects and instead build on studies 5-7. Furthermore, it will be necessary to explain WHY people are less likely to initially consider subtractive solutions and – in my opinion – it is less likely that these solutions are inherently less accessible, but rather that they are usually not perceived as desirable (even though they objectively may be).

Referee #3 (Remarks to the Author):

This paper presents seven experiments showing that people tend to overlook subtractive solutions, even if they are objectively better than alternative solutions that are additive, because subtractive solutions come to mind less readily than additive solutions. Moreover, people are more likely to overlook subtractive solutions when under cognitive load and less likely to overlook subtractive solutions when explicitly cued to consider subtraction.

I think the idea that this paper proposes has the potential to become one of the classic, fundamental insights from cognitive psychology. The finding that people tend to overlook subtractive solutions because they are less accessible than additive solutions is not only theoretically important for understanding human problem solving, but it is also practically important for improving policy making, management, design, engineering, and more.

With some refining of the theoretical development, empirical package, and writing, I think this paper could be a great fit for Nature.

Theoretical Development

- I would like to see the authors provide a more substantial theoretical explanation for why subtractive solutions are likely to be less accessible than additive ones. The current theoretical development is missing a review of past research that might provide insight into why subtractive solutions might be less accessible than additive ones. Instead, the authors rely mostly on anecdotes and controlled and uncontrolled observations, some with extremely small sample sizes.
- I recommend moving select anecdotes and observations from the introduction to the general discussion (and cutting the rest). The anecdotes and many of the observations are not sufficiently rigorous to be used as evidence with which to evaluate the hypotheses. However, some of these observations could be appropriately used to hint at possible boundary conditions and speak to potentially fruitful directions for future research.

Empirical Package

- Study 1: I recommend keeping Study 1 but acknowledging that the “control” condition contains an explicit cue to consider addition. In a creative and incentive-compatible design, this study shows that people overlook subtractive solutions, even when they are objectively better than alternative solutions that are additive. Importantly, it shows that people are more likely to think of subtractive solutions when cued to do so by pointing out that, “removing pieces is free and costs nothing.” However, I would like to see the authors acknowledge that the “control” condition contains an explicit cue to consider addition by pointing out that, “Each piece that you add costs ten cents.”
- Studies 2-4: I recommend reporting Study 4 only or Studies 3-4 separately, in greater detail. Studies 2-4 show that participants were much less likely to generate subtractive solutions than additive solutions, and importantly, in support of the relative accessibility account, that explicitly cuing participants that they could “add or subtract” increased the percentage of participants who generated at least one subtractive solution but not the percentage who generated at least one additive solution. Additionally, Study 4 shows that this pattern was robust to whether participants had a goal to “improve” or “worsen” (I was glad to see this, as I wondered whether the instruction to “improve” might bias participants toward additive solutions since these solutions allowed for greater creativity on this task.). For study 4, I would like to see the authors report not only that

explicitly cuing participants that they could “add or subtract” increased the percentage of participants who generated at least one subtractive solution, regardless of the goal, but also that explicitly cuing participants that they could “add or subtract” did not affect the percentage of participants who generated at least one additive solution, regardless of the goal. I recommend cutting Study 2 because the explicit cue did NOT significantly increase the percentage of participants who generated at least one subtractive solution in this study (interrater reliability was also a bit low in this study). Study 3 could be included or not, as the authors wish, as it provides a solid demonstration of the key findings but does not provide any additional information that is not already provided in Study 4. If Study 3 is included, be sure to correct the key to Figure 1. Consider reporting any studies that are cut in the supplemental information.

- Studies 5-7: I recommend reporting Study 5 or 6 only or Studies 5-6 separately, using a consistent analysis as in previous studies. Studies 5-7 show that people are less likely to generate subtractive solutions than additive solutions, even when the subtractive solutions are objectively better, and importantly, in support of the relative accessibility account, cognitive load decreases the likelihood that people will generate subtractive than additive solutions. I recommend cutting Study 7 because the cognitive load manipulation did NOT significantly decrease subtractive solutions (interrater reliability was also a bit low in this study). Studies 5 or 6 or both Studies 5-6 could be included, as the authors wish, as they both provide solid demonstrations of the key findings using different cognitive load manipulations (Study 5 = concurrent head-nodding, Study 6 = concurrent digit-search task), but please report them separately. Consider reporting any studies that are cut in the supplemental information. Additionally, I also recommend analyzing the results consistently across studies. Whereas Studies 2-4 focused the analysis on the likelihood that participants would generate at least one subtractive solution and Studies 5-7 could have employed the same analysis, instead, Studies 5-7 focused the analysis on the number of subtractive solutions participants generated. Please revise to be consistent or justify the decision to use different analyses across studies.

Writing

- I recommend that the authors use a more formal scientific writing style (e.g., “This paper tests whether subtractive changes come to mind less readily than additive changes”) rather than a narrative writing style (e.g., “This paper describes how we developed and tested the proposal that subtractive changes come to mind less readily than additive changes”).
- I recommend that the authors take care to describe any observations that they allude to in the discussion and specify where to find them in the supplemental information. There are currently a few instances in which the authors refer to observations without providing this information.
- I recommend that the authors more explicitly highlight the theoretical and practical contributions of their research in the general discussion.

Author Rebuttals to Initial Comments (please note that the authors have quoted the reviewers in bold font and responded in plain and italic font):

III. Referee #1:

This manuscript examines the possibility that subtractive changes to an existing artifact are less cognitively available and thus less likely to be generated or employed than additive changes. The authors show this effect is general across multiple contexts and cultures. These results are fundamental and are therefore of immediate interest to many people in my own discipline and beyond.

We really appreciate this encouraging feedback. We hoped to convey that these results are fundamental and generalizable, while also providing an account that was specified enough to make its theoretical grounding clear. Reviewer 1’s feedback helped push us even closer to

that balance, especially by urging us to provide more detail on the role of chronic versus momentary accessibility.

R1.1a. I have one primary concern and one recommendation for this manuscript. The concern is that the prevalence of this effect, though impressive, is contingent on some sort of norm or predisposition. There are certainly situations where subtractive changes are top of mind, so one question is how general the effect really is. Shown a picture of a hoarder's home, a cancerous growth, a flooded field, a sprawling suburb, etc., I believe subtractive improvements would readily come to mind for most of us. It is possible that standard and norms are an important moderator. The authors touch on this in S9 and S10 and line 88-91, but I personally am curious about the limits of the effect because it would help me understand more fully why the effect occurs.

We completely agree that certain original stimuli are likely to prompt subtractive transformations from most people. As we developed this work, we privately referred to this set of situations as “the sardine problem” because our go-to example – while less eloquent than Reviewer 1’s list – was that most people would not fail to think of subtracting a sardine from their grilled cheese sandwich. Reviewer 1’s feedback helped us to realize that we needed to be more systematic in the revision than we were in the original about what differentiates these cases. We also moved this point earlier in the manuscript so that what follows is appropriately contextualized.

We think the crux of this issue is about *what* is differentially accessible. In the revision, we are more specific that it is the mental act of searching for additive transformations versus the mental act of searching for subtractive transformations. We further specify in the revision that the phenomenon of interest is a *chronic*-accessibility effect. We claim that additive search is a chronically accessible memory-search strategy, meaning that, on average, across people and situations, it is the strategy that is more likely to be activated when people need to search for a solution. (We draw primarily from Higgins (1996) in our approach, but this is broadly consistent with many other approaches to the accessibility and use of procedural knowledge; e.g., Smith (1994) on procedural knowledge and processing strategies, Anderson (1987) on productions, which are both now cited). By sharpening our conceptualization on this point, we can now offer more satisfying answers to each of Reviewer 1’s sub-points.

- Acknowledging that chronic accessibility is not the only determinant of knowledge activation allows us to illustrate how certain stimuli could bring subtraction to ‘top of mind’ even while addition is still chronically accessible for the average participant. We are drawn to Higgins’ pithy illustration that “the high chronic accessibility of the construct ‘stick’ does not overwhelm one’s ability to identify a snake.” Similarly, we assume that a subtraction-invoking stimulus like a cancerous growth or a flooded field could prompt subtractive transformations even for people whose heuristic search strategy – had it been needed – would have been additive. We suspect that such instances of “automatic subtraction” (or addition, for that matter) may be better conceptualized as “goals” (a representation of a specific desired endstate) with clear means (remove X to produce a state with X removed) rather than “problems” (using Duncker’s classic definition, “whenever one cannot go from the given situation to the desired situation simply by action, then there has to be recourse to thinking”). For most people, a cancerous growth or a flooded field would immediately activate a representation of the desired state and the necessary action as one in the same (“get rid of X”). If for any reason that exact knowledge was not activated for a given individual and that person had

to conduct a memory search, we would predict, all things equal, that the initial memory search would be additive rather than subtractive. In short, we think that these are exceptions to *subtractive behavior* but not necessarily exceptions to the *chronic inaccessibility of subtractive search*.

(pp.7-8) *Further, to ensure that transient, task-specific influences on knowledge activation did not overwhelm and conceal the role of chronic accessibility, we used neutral prompts (e.g., “change” or “improve”) rather than connotative additive or subtractive prompts (e.g., “bolster” or “hone,” respectively), and we avoided original artifacts with obvious superfluities (e.g., a bacon, lettuce, tomato, and peanut butter sandwich) or omissions (e.g., a lettuce and tomato sandwich). Within these specifications, each of our experiments compared a condition in which people’s ideas should depend relatively more on chronic accessibility (i.e., subtraction cue not present in environment, shallower search demanded, or higher cognitive load on the participant) to a condition in which they should depend relatively less on chronic accessibility (i.e., subtraction cue present in environment, deeper search demanded, or lower cognitive load on the participant). We predicted that subtraction rates would be higher in the latter conditions.*

- The sharpened conceptualization also helps to contextualize the search for moderators related to norms or dispositions that Reviewer 1 is wondering about. While this paper focuses on *stability* in chronic accessibility (i.e., we claim that it is a general default tendency), documenting *variability* in chronic accessibility would be extremely useful, both from a practical perspective (e.g., how to mitigate the problem) and from a theoretical perspective (e.g., to pinpoint its etiology). We offer an expanded discussion of these points on pp. 14-15:

To be clear, our claim is not that chronic accessibility accounts for all systematic variation in subtractive versus additive transformations. A third contribution of this work is stimulating an agenda for future research on other determinants and moderators of subtraction neglect. With respect to the knowledge-activation stage, it is likely that stimulus information, task properties, and other incidental features of the problem context also play a role. Illustrating the non-exclusive role of chronic accessibility in knowledge activation, Higgins wrote that “chronic accessibility of the construct ‘stick’ does not overwhelm people’s ability to identify a snake”⁸; we similarly expect that people’s readiness to conduct additive searches would not overwhelm their impulse to subtract dangerous, disgusting, or dissonant components from an original (such as the grilled-cheese-with-chocolate sandwich from study S10). A key distinction may lie in whether the task at hand activates specific declarative knowledge (e.g., “anchovies are yucky”) and, in turn, a specific goal with a clear endstate and obvious means; or remains ambiguous enough to require the broader memory search associated with problem solving. Our work also invites further examination of the evaluation stage of problem solving. We focused on the prerequisite knowledge-activation phase here, but choice becomes a relevant source of variance whenever both additive and subtractive transformations have been identified. If people are also biased to prefer additive transformations, this would exacerbate subtraction neglect. Indeed, many of the candidate distal causes of an additive search strategy might be proximal contributors to an addition bias when both addition and subtraction do come under consideration.

Our findings, which are based on convenience samples of participants in the United States, also raise important questions about cultural generalizability. Preliminary studies with German and Japanese university students suggest that the chronic inaccessibility of subtraction extends beyond the United States (see study S12) but do not provide a full examination of culture as a candidate moderator. What is the role of industrialization; of resource availability; of aesthetic preferences; of social norms? Though our current research emphasizes the widespread nature of subtraction neglect, future research that explains systematic variability will be of great value in pinpointing its social, cognitive, and developmental origins; and in suggesting potential interventions to reduce its deleterious consequences.

We are excited by the idea of conducting further research to catalog situational and dispositional moderators of the chronic-accessibility effect, but we think that this work would take us beyond the scope of an already-broad first manuscript. At the same time, we hope that the extended discussion of what is known versus unknown empirically and what is expected theoretically helps to ground and define the impact of the current package.

R1.1b. Furthermore, I am not sure that many contexts would allow for balance in additive and subtractive thoughts. In a few studies (in the main text, 2-4), it is important to recognize the proximity of the lower limit on subtractive changes relative to that of the upper limit on additive changes. Yes, there are things that could be removed from the golf stimuli, but there are an infinite number of things that could be added.

Thank you for the nudge to give more attention to the theoretical and empirical implications of this point. We have incorporated this point into our expanded discussion of the socioecological aspect of the phenomenon. Like many heuristics, we suspect this search tendency emerged because it suits many transformation settings – the lower bound on subtractive transformation opportunities may be one aspect of the world that leads to chronic accessibility in the first place.

(p.6) [A]ll of these processes operate in an environment that may, from a probabilistic perspective, offer more good opportunities to add than good opportunities to subtract³⁵: new artifacts require building before honing; the number of subtractable components is always bound by what exists; and, in designed environments, one may only infrequently encounter an artifact from which others have not already removed the obviously negative components.

On the methodological side, we believe the experimental results are protected from asymmetric-bounding as a potential artifact. Though interpretation of the observational studies depends on a comparison between additive and subtractive outcomes, interpretation of the experimental studies depends on a comparison of subtractive outcomes in one condition to subtractive outcomes in another condition. Thus, any asymmetrical bounds would be controlled across conditions.

Our choice to use a relatively coarse measure of subtraction (i.e., “at least one subtractive idea”) in the miniature golf studies (experiments 2-4) also guards against this concern to some extent. That measure helps to avoid letting the unlimited number of additive ideas ‘wash out’ any meaningful variance in subtractive-idea accessibility. (More detailed discussion of this specific measure is forthcoming...)

R1.2. I personally really like this work, and see it fitting in with work on query theory by Eric Johnson and colleagues. My recommendation is related to that work. I think that the authors should consider an alternative way of analyzing their data in studies 2-4. Given the authors' goal of determining the prominence or availability of subtractive changes, I believe an analysis more similar to that of Johnson, Haubl, and Kienan's (2007) query theory might be more appropriate than the current one in studies 2-4. If I understand correctly, the authors want to illustrate the relative delay in generating subtractive changes, but they don't care about the number of changes overall, for instance. They also are curious if there is a treatment effect on this delay. However, there is additional information and an important distinction between a participant who generates her first subtractive change on her 5th and final try versus the participant who generates his first and only subtractive change on his 5th out of 30 tries. To capture that, the analysis using median ranks utilized by Johnson et al could be helpful. They calculated a standardized median rank difference (SMRD) to capture and test the relative differences in availability of certain thoughts. Applied to the present data, this statistic would be calculated for each participant as $2 * (\text{Median rank of additive changes} - \text{Median rank of subtractive changes}) / \text{number of changes suggested}$, but see the cited paper and particularly the footnote within that paper for important details.

Looking into the SMRD analysis was an extremely helpful suggestion - Thank you!

Though we share Reviewer 1's enthusiasm for the interesting connections between our work and Query Theory, it had not occurred to us to consider using their analytical strategy for experiments 2-4. Upon your suggestion, we looked very closely at their analysis strategy and the context in which it was employed to try to determine if it is a better fit for our question than the current strategy. While there are some clear similarities between the kinds of data that we collected and the data that they collected, there is a key difference that we think makes the SMRD index too conservative for our data.

Before diving into that difference, though, we note at the outset that using the SMRD measure does produce results that are consistent with our overall conclusions: In meta-analysis across the three studies, subtractive ideas tended to come earlier as indexed by the standardized median rank in the 'add-or-subtract' cue condition than in the no-cue condition, Hedges $g = .35$, $p < .001$. We now include this and individual study results as a supplemental analysis.

We are reluctant to rely on the SMRD analysis as a primary approach because it likely underestimates the magnitude of the current effect. Specifically, the standardized median rank index requires an input for the median rank of *both* types of aspects.

- In the source paper, this is a median rank for *value-increasing aspects* and a median rank for *value-decreasing aspects*.
- In our dataset, this would be a median rank for *additive transformations* and a median rank for *subtractive transformations*.

For any participant who does *not* produce both types of aspects, the analysis makes a conservative assumption: it assumes that the next thought would have produced the absent aspect.

- For example, if a participant in our study gave three additive ideas (A, A, A), calculating a SMRD is not possible without some adjustment or assumption, so the analysis assumes the fourth idea would have been subtractive (A, A, A, S).

- The consequence is that the SMRD approach recognizes no difference between hypothetical Participant 1 whose list is (A, A, A) and hypothetical Participant 2 whose list is (A, A, A, S) ... or, for that matter, between either of those hypothetical participants and hypothetical Participant 3 whose list is (A, A, A, S, S, S).

This is a pretty strong assumption about a lot of people in our study.

- In the context of the source paper, this is a reasonable approach because it affects relatively few participants. In other words, based on their task, it appears that most participants generated at least one of each kind of idea. In this context, it is conservative but its consequence is relative small.
- In our context, however, this assumption has an outsized influence because more than half of our participants do not provide any subtractive ideas at all. (And, as directly demonstrated by our analytic strategy, this is *significantly* more true of the control condition than of the prime condition.)

In our context, then, we believe this analysis is too conservative – It essentially “squashes” the very effect that is of interest here.

- To follow through with the example, we are significantly more likely to see someone like P1 (A, A, A) in our control condition and someone like P2 (A, A, A, S) or P3 (A, A, A, S, S, S) in our prime condition. This is the effect of interest that comes through with our ‘at-least-one’ approach, but it is relatively obscured using the SMRD approach.

In other words, this analytic approach is biased toward the *order* of ideas, sometimes entirely to the exclusion of the *presence* of ideas. In any case (like the source paper) where presence accounts for relatively little variance, this is extremely useful for capturing subtleties in order. However, in our case, this analysis treats a participant who thinks of zero subtraction as identical to a participant who thinks of several subtractions but does so only after at least one addition.

All of this is to say that we think the SMRD is too conservative to feature as the primary analysis; but, at the same time, it serves as a very compelling robustness check. Even in the “worst case scenario” for our prediction—a scenario in which we assume that every single participant who did not report a subtractive idea would have reported a subtractive idea if they persisted for just one more report—we still see an overall effect. In light of these considerations, we think it is best to retain the original approach in the main text and to add the SMRD approach to the Supplemental Materials as a strong robustness check. We think this combination provides a compelling empirical package and we appreciate the suggestion.

Finally, we note that this comment compelled us to bolster the connections between our work and Query Theory on the theoretical side. In the original submission, we cited Weber and Johnson’s influential “Constructing Preferences from Memory” chapter (2006) to help readers make this connection, but in the revision we have been more explicit. In addition to citing the empirical paper by Johnson and colleagues (2007), we also call attention to the theoretical connection in the discussion:

(p.14) *Though there has been less research on the chronic accessibility of procedural knowledge than of declarative knowledge¹⁴, heuristic memory-retrieval processes have been invoked to explain broad tendencies in hypothesis testing⁴¹ and evaluation¹⁶. The current work adds transformation behavior to that list.*

IV. Referee #2**(Summary comments)**

The authors propose that, when looking for improvement, people are biased toward additive changes at the expense of subtractive changes - and that this happens in part because subtractive changes come to mind less easily (are chronically less accessible).

I think the authors do a good job of demonstrating that people are indeed biased toward additive changes, but I don't think that they have demonstrated that this is due to subtractive changes being chronically less accessible.

I want to be clear: I do think that the phenomenon itself, that people are more likely to look for additive than subtractive improvements is an important and likely robust phenomenon. However, to examine it, I would suggest avoiding studies that can be explained by demand/norm effects and instead build on studies 5-7. Furthermore, it will be necessary to explain WHY people are less likely to initially consider subtractive solutions and – in my opinion – it is less likely that these solutions are inherently less accessible, but rather that they are usually not perceived as desirable (even though they objectively may be).

We thank Reviewer 2 for thoughtful feedback throughout. At a high level, this feedback really pushed us to distinguish between the different ways that preference might operate (and we hope that the response to Editor Summary comments above will also be helpful in this regard).

On the empirical front, we have taken a slightly different tack than suggested, but we are pointed at the same goal. In the revision, we have more directly addressed the norms / expectations concern in the priming studies by adding study S11 to complement experiment 1; by emphasizing some methodological details; and by adding experiment 5 as a bridge between the priming manipulations and the cognitive load manipulations. We offer details on all of these points in the Editor Summary comments above and in the specific responses to Reviewer 2 below.

On the theoretical front, we have worked to better contextualize the experimental part of the investigation relative to the observational part; to situate the empirical contribution within a broader theory; and to better ground our predictions in the indirect evidence that is available in the literature. We offer additional details and point to excerpts from the revision in the Editor Summary comments above and in the specific responses to Reviewer 2 below.

(Motivating observations comments)

The first set of “motivating observations,” which are just intended to demonstrate the phenomenon rather than nail down the process, actually illustrate some of the many reasons why people are biased toward addition. That is, they can all be explained without having to invoke a novel explanation based on differences in chronic accessibility. For instance, the fact that members of a university community propose far more additive than subtractive changes when queried by an incoming president can be explained by (1) political/social reasons: it can be problematic suggesting shutting down an initiative championed by someone else, (2) assumptions that “if it is there, it is there for a reason,” (3) sunk cost bias (we already invested so much in these programs), (4)

wanting to be recognized for creativity and effort (proposing a new program signals more creativity than proposing shutting one down), etc... The main reason, though, which I think applies to every example, is that people assume they are SUPPOSED to add: the president is asking them to come up with new ideas, not shoot down old ones. In other words, the bias toward additive changes can be accounted for by norms and expectations.

We are grateful that Reviewer 2 took the time to spell out these possibilities in detail. Even in the initial submission, we worked hard to design our experiments in ways that would allow us to rule out preference in favor of accessibility – but we did not give as much explicit attention to *expectations* as Reviewer 2's comments convinced us that we should have. This important issue motivated us to design and run a new supplemental experiment (S11), to introduce a new experiment (experiment 5), and to emphasize some methodological details in the original experiments that should also help ease some of these concerns.

Before describing these experimental details below, we note that the observational studies have a more general goal than the experiments. We thank the Reviewer for bringing this lack of clarity to our attention, and we now make it clear that we relied on the observations to illustrate the breadth of relevance of subtractive versus additive transformations and to determine whether there is a broad behavioral tendency to investigate in the first place. We agree that the behavioral outcomes in the observational studies are likely multiply determined and have been even more explicit about that in the revision.

(p.5) While the strong tendencies documented in the observational studies are probably multiply determined, we focused our subsequent theorizing and experimentation on the idea-generation phase because it precedes evaluation. Preference-based explanations are relevant only when people consider both additive and subtractive actions. We propose here that people are not equally likely to consider subtraction in the first place.

(p.15) Our work also invites further examination of the evaluation stage of problem solving. We focused on the prerequisite knowledge-activation phase here, but choice becomes a relevant source of variance whenever both additive and subtractive transformations have been identified. If people are also biased to prefer additive transformations, this would exacerbate subtraction neglect. Indeed, many of the candidate distal causes of an additive search strategy might be proximal contributors to an addition bias when both addition and subtraction do come under consideration.

(Experiments comments)

The second set of studies, the actual experiments, aim to go beyond demonstrating the phenomenon (which seems real but also multiply determined) and showing the process (subtractive solutions being less accessible).

Experiments 1-4 do this by demonstrating that people are more likely to offer subtractive changes if they are explicitly told (cued) that this is an option. This doesn't demonstrate the role of chronic accessibility as it is fully consistent with the norms explanation (or, in experimental terms: the cue manipulation activates a demand/conversational norms effect: it tells participants that subtractive changes are allowed, something many may not have assumed otherwise).

Experiments 5-7 examine the mechanism by showing that subtractive solutions are even less likely when people are under cognitive load. I like this approach more than that of exps 1-4 as it doesn't manipulate norms or expectations (the way the cuing manipulation did). The problem with these studies, however, is that the results are much weaker than for the previous studies. My take-away is that these studies do illustrate that subtractive solutions are less likely to come to mind initially than additive ones, but that this only contributes minimally to the (rather robust) phenomenon that people are more likely to offer additive solutions.

Again, thank you for this feedback, which we believe prompted substantial empirical improvement in the revision. It also led us to emphasize some under-specified details of the original experiments. We will consider the claim for each of the two paradigms.

First, this feedback led us to ask ourselves whether the cue to consider subtraction in the experiment 1 paradigm (transform a Lego structure to hold a brick over a figurine) could have changed participants' evaluation of a subtractive approach. We think a fair way to characterize the concern here is:

Were participants in the control (vs. subtraction-cue) condition equally likely to *think* of the subtractive transformation; but more likely to *reject* subtractive transformations because they thought it might not be allowed? In other words, did the subtraction cue give participants permission to use subtraction?

- We first note one detail about our method that tempers our concern about this alternative explanation. Instructions were explicit that “You may alter the structure however you want.” The instructions were very brief, so it is unlikely that this was lost in the fine print, and they were delivered to participants both in writing and verbally by the experimenter. We regret that this detail was buried in the Supplemental Material of the original submission and have moved it to the main text given its importance.

Of course, even with the ‘blanket permission,’ it is true that participants in the subtraction cue condition had direct explicit permission that they could add or subtract whereas participants in the no-cue condition only had direct explicit permission that they could add. This led us to design study S11.

- This study addresses the alternative explanation by posing the following question: when participants do consider both subtraction and addition, does the set of instructions in the subtraction cue (vs. control) condition make them more likely to view subtraction as legitimate and/or more likely to choose it? Another way to think of this experiment is that we took knowledge activation out of the equation (by showing participants both transformations) and observed whether the cue manipulation changed their evaluation of the transformations. After participants read one version of the instructions (subtraction cue vs. control), they viewed gifs of the additive and subtractive transformations. We asked them to indicate if each transformation was allowed by the instructions and to choose which one they thought was best.

(pp.61-62, Supplemental Material 2.11) *We designed this experiment as a post-hoc assessment of a potential alternative explanation for the results of experiment 1. In experiment 1, participants were more likely to transform a Lego tower by subtraction when instructions included an explicit subtraction cue than when they did not. Noting that instructions in both conditions gave participants blanket permission to alter the structure in any way, we concluded that the subtraction cue increased subtractive transformations because it activated the idea. However, a potential alternative interpretation is that the subtraction cue increased subtractive transformations because it increased their perceived legitimacy. In other words, it is possible that a non-trivial number of participants in the control condition thought of subtraction, but passed it up in favor of an additive approach because they inferred that subtraction was not allowed; whereas participants in the subtraction-cue condition had been assured that subtraction was allowed because it was mentioned explicitly. To investigate this alternative possibility, we designed a parallel task that held knowledge activation constant (by presenting both transformation possibilities) and measured participants' evaluations of each kind of transformation following one of the two instruction sets. If the difference in experiment 1 is attributable to participants' evaluation of the solutions, rather than to the differential activation of subtractive ideas, then we would expect the instructions to affect evaluations in the parallel set-up of study S11.*

Because the goal of the experiment was to determine whether the subtraction-cue (vs. control) instructions legitimized the subtractive transformation in participants' eyes, we needed a paradigm that would allow us to present both kinds of transformations to participants without seeming to provide the researchers' endorsement of either one. We created a minimal managerial scenario, telling participants that they were the manager and that their team had come up with "two possible ideas" in response to the task specifications from experiment 1. Using this approach, we could deliver the same instructions, show the same Lego setup, offer the same contingent bonus, and solicit participants' evaluations of both the additive and subtractive transformations, without representing either approach as approved or preferred by the researchers....

Results and interpretation are summarized in the main text:

(p.9) Reinforcing the knowledge-activation interpretation over the preference interpretation, we found that the same cue manipulation did not significantly affect whether these participants viewed the subtractive transformation as legitimate (control = 98%, cue = 100%), nor did it affect their likelihood of choosing the subtractive transformation over the additive transformation in a head-to-head choice (control = 82%, cue = 86%, $\chi^2 = 0.47$, $p = .491$, $\phi = .05$).

We take this as compelling evidence for the accessibility explanation over the alternative permission-based preference explanation.

We think a preference explanation for the mini-golf task is less likely for two main reasons. First, the dependent variable is different. In contrast to the Lego paradigm, it

is possible that expectations are relevant to the mini-golf paradigm – and specifically, the expectations of the “client” rather than the researcher (“What does Mr. Popo want me to do?”). However, it is important to recall that we asked participants to list *all* of their ideas (up to 24) rather than just producing a single transformation, which should limit the role of preference. Second, it is unlikely that the same norms or expectations would apply to making things worse as to make things better, and yet we replicated the effect in the make-it-worse condition of experiment 4.

(pp.9-10) Whereas experiment 1 potentially invited both activation and evaluation processes by soliciting only a single transformation, experiments 2 and 3 asked participants to list “all of the different ways that [they] might be able to improve” the original. This thought-listing measure minimizes the potential role of evaluation and offers a richer view of relative accessibility^{33,37}. Ensuring that the conditions did not asymmetrically offer tacit endorsement of either kind of transformation, the no-cue instructions did not mention either addition or subtraction, whereas the cue instructions mentioned both, reminding participants that they could “add or subtract.”

Even with study S11 in hand, a remaining concern might be that any introduction of the subtraction idea (whether in instructions or presented as a transformation that someone produced) has to come from somewhere. Barring a mechanism to directly activate subtraction in someone’s brain, we cannot 100% rule out some social interpretation of any prime. Based on this recognition, we think that the new experiment 5 is a strong complement to study S11. The approach of experiment 5 was to give participants more chances to activate a subtractive transformation on their own, without introducing any information “from the outside.”

(p.11) In contrast to experiments 1–4, which directly manipulated whether subtractive ideas came to mind quickly and easily, experiments 5–8 sought to manipulate whether participants would get beyond a shallow search and potentially activate subtractive ideas on their own. In experiment 5 (N = 299), we presented participants with a digital 10 × 10 grid of white and green boxes, a novel stimulus in which the individual components have no inherent value. Participants could toggle the color of any box by clicking on it. They read that their goal was to change the original grid to make it symmetrical left-to-right and top-to-bottom using the fewest clicks possible. Figure 2 illustrates that the original grid had extraneous filled boxes in one corner of the grid. Participants could achieve symmetry by adding to the three empty corners of the grid or by subtracting from the marked corner of the grid. Unlike the grid task in study S1, this version had an objectively correct transformation. Of course, participants could only recognize the subtractive transformation as correct if they thought of it.

In the control condition, participants proceeded immediately to the critical trial. In the repeated search condition, participants first completed practice trials on three structurally similar grids. Crucially, participants in this condition received no external feedback on their practice responses. The repetition condition merely gave them more opportunities to get beyond the chronically accessible approach. Increasing the cumulative probability of an incidental subtraction discovery during practice should increase one’s likelihood of employing a subtractive search in the critical trial. Consistent

with the chronic-accessibility hypothesis, 49% of participants produced the subtractive solution in the control condition, whereas 63% produced it in the repeated search condition, $\chi^2 = 5.87$, $p = .015$, $\phi = 0.15$.

To summarize, we hope that the combination of methodological details, generalizability across goals, and new studies eases concerns about the role of preference in the documented effects. With three experimental approaches (direct activation manipulations, the repeated search manipulation, and cognitive load manipulations), we think this is compelling evidence for the chronic-accessibility effect.

(Conceptual comments)

Conceptually, I was surprised that the authors do not offer an explanation for why subtractive solutions are less likely to come to mind. The authors explicitly separate the accessibility (consideration) explanation from the fact that people may PREFER additive changes for a variety of reasons (e.g., see the reasons I specified above for the additive university improvements). However, rather than assuming that subtractive changes are inherently less accessible (for an unknown reason), I think it is more likely that people are simply less likely to first go that route (i.e., consider them) because they usually do not want to. That is, preference drives consideration (as it usually does – it is odd to separate them). In fact, I would argue that subtractive changes are actually easier to imagine than additive ones – as it doesn't require any creative search. This is one of the reasons why people are more sensitive to errors of commission than to errors of omission: it's much easier for us to generate the counterfactual that something had not happened than to generate the counterfactual that something did happen.

As also discussed in our overview comments to the Editor, this comment pushed us to be much more careful in how we addressed the relationship of preference and accessibility, not only from an empirical point of view, but also from a theoretical point of view. In short, we agree that “preference drives consideration” in the sense that it may contribute to the chronic accessibility ‘in the first place’; at the same time, we think that our experimental package provides compelling evidence that chronic accessibility is driving the observed results more so than preference.

Based on this comment we were also encouraged to move and expand our discussion of the potential distal reasons for chronic accessibility from a brief comment in the discussion to a full paragraph in the introduction. It is indeed the case that we were primarily motivated from an observational / empirical perspective. Given that addition was so prevalent in the observational studies, we felt there was good reason to consider why people were not subtracting. As we explain in the manuscript, idea generation and evaluation are candidate responses, but (p.5) “*we focused our subsequent theorizing and experimentation on the idea-generation phase because it precedes evaluation. Preference-based explanations are relevant only when people consider both additive and subtractive actions.*”

Though we favored this prediction based on the possibility suggested primarily by our own data, we also draw attention to a number of findings from the literature that indirectly or circumstantially supported it:

(p.6) Though previous theorizing has not conceptualized subtractive versus additive transformations, there are cognitive, cultural, and socioecological reasons to suspect that over many transformation opportunities, people might rely on additive

transformations more routinely than subtractive transformations, to the point that they become a chronically accessible target of mental searches. First, with respect to perception and categorization, additive transformations may be incrementally easier to process. Any subtractable component must first be understood as part of the artifact before it can potentially and effortfully be considered as not part of the artifact²⁶. Addable components do not require such recategorization. Second, over time, additive searches may become more positively valenced than subtractive searches. Numerical concepts of 'more' and 'higher' may map to evaluative concepts of 'positive' and 'better'^{27,28}; tangible contributions are culturally valued^{29,30}; and acquiring and displaying resources is fitness enhancing³¹. Third, people might be reluctant to subtract because of attentional and evaluative processes that favor the status quo³²⁻³⁴. Finally, all of these processes operate in an environment that may, from a probabilistic perspective, offer more good opportunities to add than good opportunities to subtract³⁵: new artifacts require building before honing; the number of subtractable components is always bound by what exists; and, in designed environments, one may only infrequently encounter an artifact from which others have not already removed the obviously negative components.

We think the reviewer's feedback also highlights the likelihood of a self-reinforcing cycle of some kind. If people do not prefer subtraction, then subtraction will not come to mind, then people will learn to prefer addition (not subtraction) even more, then subtraction will not come to mind, and so on.

(p.15) Our work also invites further examination of the evaluation stage of problem solving. We focused on the prerequisite knowledge-activation phase here, but choice becomes a relevant source of variance whenever both additive and subtractive transformations have been identified. If people are also biased to prefer additive transformations, this would exacerbate subtraction neglect. Indeed, many of the candidate distal causes of an additive search strategy might be proximal contributors to an addition bias when both addition and subtraction do come under consideration.

(p.14) For anyone who aims to unlock the cache of subtractive ideas, it is important to further recognize that the chronic inaccessibility of subtraction is likely to be a self-reinforcing phenomenon. Each overlooked subtraction would lead to more practice with non-subtractive strategies. More practice would create stronger associations and habits, further disadvantaging the accessibility of subtraction. Without effective intervention, then, additive search is likely to continue to predominate.

V. Referee #3

This paper presents seven experiments showing that people tend to overlook subtractive solutions, even if they are objectively better than alternative solutions that are additive, because subtractive solutions come to mind less readily than additive solutions. Moreover, people are more likely to overlook subtractive solutions when under cognitive load and less likely to overlook subtractive solutions when explicitly cued to consider subtraction.

I think the idea that this paper proposes has the potential to become one of the classic, fundamental insights from cognitive psychology. The finding that people tend to overlook subtractive solutions because they are less accessible than additive solutions is not only theoretically important for understanding human problem solving, but it is also practically important for improving policy making, management, design, engineering, and more.

We are honored by the suggestion that this could become a classic, fundamental insight. We are thus all the more motivated to provide the clearest account that we can at this stage. We appreciate the suggestions below that helped to move us in that direction.

With some refining of the theoretical development, empirical package, and writing, I think this paper could be a great fit for Nature.

R3.1 - Theoretical Development

- I would like to see the authors provide a more substantial theoretical explanation for why subtractive solutions are likely to be less accessible than additive ones. The current theoretical development is missing a review of past research that might provide insight into why subtractive solutions might be less accessible than additive ones. Instead, the authors rely mostly on anecdotes and controlled and uncontrolled observations, some with extremely small sample sizes.

- I recommend moving select anecdotes and observations from the introduction to the general discussion (and cutting the rest). The anecdotes and many of the observations are not sufficiently rigorous to be used as evidence with which to evaluate the hypotheses. However, some of these observations could be appropriately used to hint at possible boundary conditions and speak to potentially fruitful directions for future research.

Thank you for these set-up suggestions. Given that our full package took a less traditional approach by starting with an open exploratory question and observational studies, we moved too quickly in the initial submission from introducing a hypothesis that might (partially) account for the observations directly to testing that hypothesis. Your comment and other reviewer comments reminded us that readers would greatly value a broader review of any literature that could help them assess the theoretical plausibility of the specific predictions. As detailed above, we have moved our discussion of the possible reasons for the chronic accessibility effect from a parenthetical in the discussion to a full paragraph in the hypothesis development section.

(p.6) ... [T]here are cognitive, cultural, and socioecological reasons to suspect that over many transformation opportunities, people might rely on additive transformations more routinely than subtractive transformations, to the point that they become a chronically accessible target of mental searches...

In the revision we have also tried to be clearer about the separate stages that this work covers, with anecdotes motivating observations, observations and suggestive literature motivating causal claims, and clear theoretical hypotheses motivating experimental design.

We appreciate the review team's openness to a less traditional narrative structure, and we are grateful for suggestions that helped us try to maximize the value of a combined exploratory and hypothesis-driven approach in one manuscript.

Empirical Package

R3.3. Study 1: I recommend keeping Study 1 but acknowledging that the “control” condition contains an explicit cue to consider addition. In a creative and incentive-compatible design, this study shows that people overlook subtractive solutions, even when they are objectively better than alternative solutions that are additive. Importantly, it shows that people are more likely to think of subtractive solutions when cued to do so by pointing out that, “removing pieces is free and costs nothing.” However, I would like to see the authors acknowledge that the “control” condition contains an explicit cue to consider addition by pointing out that, “Each piece that you add costs ten cents.”

Thank you for this recommendation. We did not mean to sacrifice precision for brevity in the original, but now recognize why it needs to be more transparent. We have added more complete description for both conditions and revised the condition names to avoid any confusion. Revisions to the supplementals are consistent with these changes in the main text:

(p.8) In the control condition, instructions explicitly mentioned addition (“each piece that you add costs ten cents”) but not subtraction. In the subtraction-cue condition, instructions also mentioned subtraction (“each piece that you add costs ten cents but removing pieces is free”). Instructions in both conditions gave participants blanket permission to “alter the structure however you want.”

We also acknowledge this difference in the priming manipulations as one of the differences between experiment 1 and subsequent priming experiments.

(pp.9-10) Ensuring that the conditions did not asymmetrically offer tacit endorsement of either kind of transformation, the no-cue instructions did not mention either addition or subtraction, whereas the cue instructions mentioned both, reminding participants that they could “add or subtract.”

3.4. Studies 2-4: I recommend reporting Study 4 only or Studies 3-4 separately, in greater detail. Studies 2-4 show that participants were much less likely to generate subtractive solutions than additive solutions, and importantly, in support of the relative accessibility account, that explicitly cuing participants that they could “add or subtract” increased the percentage of participants who generated at least one subtractive solution but not the percentage who generated at least one additive solution. Additionally, Study 4 shows that this pattern was robust to whether participants had a goal to “improve” or “worsen” (I was glad to see this, as I wondered whether the instruction to “improve” might bias participants toward additive solutions since these solutions allowed for greater creativity on this task.). For study 4, I would like to see the authors report not only that explicitly cuing participants that they could “add or subtract” increased the percentage of participants who generated at least one subtractive solution, regardless of the goal, but also that explicitly cuing participants that they could “add or subtract” did not affect the percentage of participants who generated at least one additive solution, regardless of the goal. I recommend cutting Study 2 because the explicit cue did NOT significantly increase the percentage of participants who generated at least one

subtractive solution in this study (interrater reliability was also a bit low in this study). Study 3 could be included or not, as the authors wish, as it provides a solid demonstration of the key findings but does not provide any additional information that is not already provided in Study 4. If Study 3 is included, be sure to correct the key to Figure 1. Consider reporting any studies that are cut in the supplemental information.

We appreciate this added perspective on how to most clearly convey our results to readers.

We have added the requested result for experiment 4, which shows that cuing participants to “add or subtract” did not affect the percentage of participants who generated at least one additive solution. At this stage we are directing readers to the analyses in the supplemental material because testing the 2×2 requires a different approach than we currently have in the main text. We thought this was useful for brevity but are very open to moving the specific statistics to the main text.

(p.11) *The cue did not affect the percentage of participants who generated at least one additive idea (see Supplemental Material section 1.4).*

(p.38 – Supplemental Material) *A logistic regression also modeled the likelihood of addition as a function of the cue condition (cue/no cue), the goal condition (make-better/make-worse), and their interaction. The regression revealed no effects of either the cue ($b = .31$, $Z = 1.04$, $p = .300$), the goal condition ($b = -.41$, $Z = -1.36$, $p = .174$), or their interaction ($b = .99$, $Z = 1.63$, $p = .103$).*

At the Editor’s direction, we have retained all studies for the time being. We look forward to working with the editorial team on necessary changes in future steps.

Finally, thank you for bringing the most-problematic typo in the Figure 1 key to our attention. We have fixed it.

R3.5. Studies 5-7: I recommend reporting Study 5 or 6 only or Studies 5-6 separately, using a consistent analysis as in previous studies. Studies 5-7 show that people are less likely to generate subtractive solutions than additive solutions, even when the subtractive solutions are objectively better, and importantly, in support of the relative accessibility account, cognitive load decreases the likelihood that people will generate subtractive than additive solutions. I recommend cutting Study 7 because the cognitive load manipulation did NOT significantly decrease subtractive solutions (interrater reliability was also a bit low in this study). Studies 5 or 6 or both Studies 5-6 could be included, as the authors wish, as they both provide solid demonstrations of the key findings using different cognitive load manipulations (Study 5 = concurrent head-nodding, Study 6 = concurrent digit-search task), but please report them separately. Consider reporting any studies that are cut in the supplemental information. Additionally, I also recommend analyzing the results consistently across studies. Whereas Studies 2-4 focused the analysis on the likelihood that participants would generate at least one subtractive solution and Studies 5-7 could have employed the same analysis, instead, Studies 5-7 focused the analysis on the number of subtractive solutions participants generated. Please revise to be consistent or justify the decision to use different analyses across studies.

Thank you for the opportunity to more clearly explain the shift in our analytic strategy across these sets of studies, which is due to the difference in DVs across sets.

First, we note here that the conclusions seem to be robust to many different analytical approaches. For example, if we collapsed the multiple-trial grid studies (previously experiments 5—7, now experiments 6—8) into a categorical “at least one” approach, results support the same conclusion:

	Control % Sub at least once	CogLoad % Sub at least once	Chi Sqr	p-value
Study 6	75%	64%	4.50	.034
Study 7	85%	73%	5.75	.017
Study 8	77%	73%	1.13	.287
Meta-Analysis:	OR = -.32, $z = -2.36$, $p = .018$			

We have added this analytic approach in the Supplemental Material.

Below, we offer our rationale for using the different approaches in the main text:

- We think that the “at least one” categorical analysis, though relatively coarse, is the simplest and most appropriate approach for the thought-listing studies (experiments 2—4) because it provides a straightforward intuition for whether a given participant thought of any subtractive transformations (assuming they followed the directions to report all of their ideas). At Reviewer 1’s suggestion, we now also report analyses using the standardized mean rank difference approach based on Johnson and colleagues’ work in the supplemental materials.
- We see the multiple-trial grid tasks as fundamentally different from the thought-listing task because each of the four answers in the grid task is tied to a different original stimulus. This gives us a 1:1 mapping for the question “did Participant X think of the subtractive solution for Puzzle I?” We have four responses for each participant (rather than the as-many-as-you-want approach for thought-listing) and so we can interpret the basic arithmetic mean and standard deviation in a straightforward way. We think this provides a good balance of simplicity, power, and precision. With that said, these studies are also robust to different analytical approaches and we now note in the Supplementals that the “at-least-one” approach yields a similar conclusion.

At the Editor’s suggestion we have retained all of the studies for now. We have also retained the meta-analytic approach in the main text, but we are certainly open to moving the separate study results from the supplemental material to the main text as directed. We are also open to adding justification for each of those approaches to the main text.

R3.6. I recommend that the authors use a more formal scientific writing style (e.g., “This paper tests whether subtractive changes come to mind less readily than additive changes”) rather than a narrative writing style (e.g., “This paper describes how we developed and tested the proposal that subtractive changes come to mind less readily than additive changes”).

Thank you for this suggestion. We have made changes in the spirit of this comment throughout the main text (including the example you mentioned) and hope that we will have

an opportunity to work on this with the editorial team if this manuscript moves further through the process.

R3.7. I recommend that the authors take care to describe any observations that they allude to in the discussion and specify where to find them in the supplemental information. There are currently a few instances in which the authors refer to observations without providing this information.

Thanks for alerting us to potential confusion here. We have now labeled each of the referenced observational studies in the Main Text so readers can more easily find their way to details in the Supplementals.

R3.8. I recommend that the authors more explicitly highlight the theoretical and practical contributions of their research in the general discussion.

Thank you for this suggestion. We have discussed this important issue in great depth in sections above, detailing how we clarified the theoretical grounding and implications of this work. We appreciate the invitation to describe in a straightforward way why we think this work can be impactful.

(p.13) Changing the physical, intellectual, and social world into something better is among the most essential work of human behavior. The primary conceptual contribution of the current research is introducing the fundamental distinction between pursuing those changes through subtraction or through addition. Asking when people change their worlds through subtraction builds a conceptual bridge between what might otherwise be treated as disparate phenomena from areas such as design and problem solving (changing objects); reasoning, learning, and communication (changing ideas); and coordination, decision making, and motivation (changing situations).

(p.13-14) The primary empirical contribution of the current research is identifying the chronic inaccessibility of subtractive searches as one of the drivers of subtraction neglect. Our experimental results show that the likelihood of subtractive transformations depends on the presence of cues that prompt subtractive search (experiments 1–4), on the number of opportunities one has to deviate from what is chronically accessible (experiment 5), and on the situational availability of cognitive resources (experiments 6–8). Though there has been less research on the chronic accessibility of procedural knowledge than that of declarative knowledge¹⁴, heuristic memory-retrieval processes have been invoked to explain broad tendencies in hypothesis testing⁴¹ and evaluation¹⁶. The current work adds transformation behavior to that list.

(p.14) From a practical perspective, the ever-growing number of frameworks that support the application of behavioral science to public and private life could valuably recognize subtraction neglect as an important factor in the design of nudges⁴²; wise interventions⁴³; and other programmatic efforts to train engineers⁴⁴, educators⁴⁵, and other innovators⁴⁶. For anyone who aims to unlock the cache of subtractive ideas, it is important to further recognize that the chronic inaccessibility of subtraction is likely to be a self-reinforcing phenomenon. Each overlooked subtraction would lead to more practice with non-subtractive strategies. More practice would create stronger associations and habits, further disadvantaging the

accessibility of subtraction. Without effective intervention, then, additive search is likely to continue to predominate.

(p.14-15) To be clear, our claim is not that chronic accessibility accounts for all systematic variation in subtractive versus additive transformations. A third contribution of this work is stimulating an agenda for future research on other determinants and moderators of subtraction neglect. With respect to the knowledge-activation stage, it is likely that stimulus information, task properties, and other incidental features of the problem context also play a role... Our work also invites further examination of the evaluation stage of problem solving...

(p.16)...The tendency to overlook subtraction may be implicated in a variety of costly trends including overburdened minds and schedules¹⁰, increasing red tape in institutions¹¹, and humanity's encroachment on the safe operating conditions for life on earth^{12,13}. If people systematically pursue adequate additive transformations before they even consider comparable, sometimes superior, subtractive transformations, they may be missing opportunities to make their lives more fulfilling, their institutions more effective, and their planet more livable.

Reviewer Reports on the First Revision:

Referee #1 (Remarks to the Author):

This manuscript examines the possibility that subtractive changes to an existing artifact are less cognitively available and thus less likely to be generated or employed than additive changes. The authors show this effect is general across multiple contexts and cultures. Like the previous round, I believe that results are fundamental and are therefore of immediate interest to many people in my own discipline and beyond.

In the previous round, I had one primary concern and one recommendation for this manuscript. The concern was about the limits of the effect. The recommendation was about an alternative analysis. The authors have been remarkably responsive to both. I whole-heartedly agree that the SMRD measure has its problems, and that it is conservative. I think the addition of the analyses using the measure is useful, nevertheless, and am glad to see it included.

I also appreciate the further attention provided to the possibility of spontaneous subtractive thinking. The revised manuscript provides, in my opinion, a more complete picture of the effect.

Referee #2 (Remarks to the Author):

I think this revision is a substantial improvement over the previous version (and I want to reiterate that I was already convinced of the importance and robustness of the basic phenomenon – I was just skeptical of the proposed explanation and the empirical support for it). There are some lingering, non-trivial issues that I would like the authors to address, but I think they can be handled with changes in wording and adding nuance to claims.

First, let me elaborate on the ways in which the manuscript has improved in my opinion:

- The authors now offer a more detailed discussion of why subtractive improvements may be less accessible than additive ones, including pointing out that they may have been more valued and rewarded in the past.
- The motivating observations are more concisely presented and more clearly separated from the actual studies that aim to demonstrate and examine the phenomenon.

- The new study 5 is a nice demonstration that people can be made to consider subtractive solutions with additional practice, without having to explicitly guide them to doing so (thus avoiding the demand effects that may be present in studies 1-4).

That being said, I think there are a few ways in which I think this paper still falls short of its potential – and that can fortunately be rather easily remedied:

The conceptualization (in both the paper and the revision notes) of problem solving as a 2-stage process consisting of a consideration stage followed by an evaluation stage, with preference (and possibly expectations) only affecting the latter stage is both overly simplistic and logically inconsistent with other parts of the paper (see the first point I listed above). Take the sentence (p. 5): “Preference-based explanations are relevant only when people consider both additive and subtractive actions.” That is simply not true. If additive strategies are more desirable (e.g., more rewarding) or more normative (e.g., what’s generally expected in these situations), then that will affect the generation of ideas, not just their evaluation. I urge the authors to remove this sentence and any other suggestion that preference or expectations only affect the evaluation stage.

In my previous review, I pointed out that the results of experiments 1-4 could be driven by participants’ expectations of what they are supposed to do. Some participants may implicitly expect that they’re supposed to add something, so those may only consider subtractive solutions when the instructions explicitly tell them it’s ok to do so. The authors now counter this with two arguments, neither of which is convincing to me. First, they state that the instructions made clear they could change it in any way (“alter it any way you want”). That of course doesn’t preclude a segment of participants to still assume that they’re supposed to add something. Second, they present a new study (S11) in which participants are explicitly presented with both subtractive and additive solutions and both types of solutions are judged as legitimate by participants, irrespective of instructions. However, explicitly presenting a subtractive solution by itself signals to participants that this is an acceptable way to approach the problem, so this doesn’t address the issue at all. This goes back to my previous point: the authors seem to assume that expectations can only affect evaluation, not idea generation. That is obviously not correct: If I implicitly assume that the experimenter is looking for additive solutions, then I will only generate additive solutions unless it is made explicit that subtractive ones should also be considered (either through instructions or by showing me subtractive solutions). I would encourage the authors to downplay study 1-4, to acknowledge that participants’ expectations may have influenced the results, and not to present S11 as a rebuttal of the role of expectations (given that showing the subtractive solutions by itself changes expectations).

I think the new study 5 is a good addition to the paper (as I noted earlier), but even though it is much subtler than studies 1-4, it is still possible that this manipulation increases subtractive solutions because it changes expectations. People who implicitly assume that they are supposed to offer additive solutions may change this assumption when they notice, in the practice trials, that each of the problem sets can be solved by changing a single constellation of squares. That is subtle, and unlikely to be a demand effect, but it may still be a way to challenge expectations. To be clear, I still think this is an interesting result. Moreover, participants’ natural assumption that they are supposed to add rather than subtract is related to the central thesis of this paper. However, rather than positioning this study as showing that people only consider subtractive solutions when they dig deeper, I would recommend that the authors acknowledge that it may also lead them to question assumptions.

In the end, I’m wondering if “chronic accessibility” is the correct term for what the authors are proposing as the mechanism driving the reduced likelihood to generate subtractive solutions. Reduced accessibility suggests that it is harder to come up with (i.e., it is harder to access in memory). Indeed, the authors now state, as one of the possible drivers of this reduced accessibility, that “additive transformations may be incrementally easier to process.” However, I

don't think that's generally true. As I mentioned in my previous review, we do know that it is easier to generate counterfactuals that undo an action rather than add one. It seems plausible to me that it is easier to simply remove something that is already there than to come up with something new to be added. Moreover, I don't think that the "accessibility" terminology is essential for the authors' account. The abstract talks about "heuristic search strategies" and I think that is a more accurate (less speculative) way to describe it: people somehow default to additive solutions. These solutions have a privileged status and are more likely to be spontaneously considered. It is about WHICH solutions you try to access rather than how EASY it is to access them.

Referee #3 (Remarks to the Author):

Thank you for your conscientious revision – the improved theoretical grounding, clarified conceptualization, added experiments, and more direct discussion of implications have addressed my concerns and enhanced the potential contribution of this work. Nice work!

Author Rebuttals to First Revision (please note that the authors have quoted the reviewers in plain font and responded in italic font):

Referee #1 (Remarks to the Author):

This manuscript examines the possibility that subtractive changes to an existing artifact are less cognitively available and thus less likely to be generated or employed than additive changes. The authors show this effect is general across multiple contexts and cultures. Like the previous round, I believe that results are fundamental and are therefore of immediate interest to many people in my own discipline and beyond.

In the previous round, I had one primary concern and one recommendation for this manuscript. The concern was about the limits of the effect. The recommendation was about an alternative analysis. The authors have been remarkably responsive to both. I wholeheartedly agree that the SMRD measure has its problems, and that it is conservative. I think the addition of the analyses using the measure is useful, nevertheless, and am glad to see it included.

I also appreciate the further attention provided to the possibility of spontaneous subtractive thinking. The revised manuscript provides, in my opinion, a more complete picture of the effect.

Thank you for your invaluable time and thought, which we agree have allowed the manuscript to provide a more complete picture of the effect.

Referee #2 (Remarks to the Author):

I think this revision is a substantial improvement over the previous version (and I want to reiterate that I was already convinced of the importance and robustness of the basic phenomenon – I was just skeptical of the proposed explanation and the empirical support for

it). There are some lingering, non-trivial issues that I would like the authors to address, but I think they can be handled with changes in wording and adding nuance to claims.

First, let me elaborate on the ways in which the manuscript has improved in my opinion:

- The authors now offer a more detailed discussion of why subtractive improvements may be less accessible than additive ones, including pointing out that they may have been more valued and rewarded in the past.
- The motivating observations are more concisely presented and more clearly separated from the actual studies that aim to demonstrate and examine the phenomenon.
- The new study 5 is a nice demonstration that people can be made to consider subtractive solutions with additional practice, without having to explicitly guide them to doing so (thus avoiding the demand effects that may be present in studies 1-4).

We are pleased that you agree these are improvements. We very much appreciate your role in helping us make them, and we appreciate your discussion of both strengths and weaknesses.

[nb. We have moved Reviewer 2's summary comment to the top because it is our general response to this comment that frames the more specific suggestions.] In the end, I'm wondering if "chronic accessibility" is the correct term for what the authors are proposing as the mechanism driving the reduced likelihood to generate subtractive solutions. ... I don't think that the "accessibility" terminology is essential for the authors' account. The abstract talks about "heuristic search strategies" and I think that is a more accurate (less speculative) way to describe it: people somehow default to additive solutions.

In broad terms, these comments have persuaded us that it makes sense to stay at the more general level when we describe the effect, referring to it throughout as a *heuristic search strategy*. This simplifies our previously confusing discussion of a distal versus proximal role of preferences / expectations and allows us to say, a bit more agnostically, that people tend to use this strategy. (We *do* think that repeated use of the strategy makes it a chronically accessible knowledge structure... but we agree that delving into this point is not essential, difficult to do with clarity in limited space, and that we can describe the key findings without it.) Importantly, individuals could have developed the strategy over time for many reasons—including preferences—and for any given judgment they might move to a different strategy if cues change their expectations about what is reasonable (and if they are willing and able to put it in the additional work).

Here are three key examples of the more general account, and our general avoidance of finer-grained distinctions about accessibility:

(Line 91) "Thus, if additive search is a common default, people should be more likely to rely on it when they are cognitively loaded; and, conversely, they should be less likely to rely on it when task experience or task information cues them to use another strategy."

**(Line 113 – Note the term *prioritizes*, which would accommodate expectations.)
"Explicitly priming a specific, relevant action category should offset the influence of a heuristic that prioritizes a different category"^{6,27,29,30}**

(Line 200) “Heuristic search surely does not account for all variation in subtractive and additive transformations. ... Task-specific features might also cue a tailored memory search (e.g., an editor’s instructions to “shorten”). Further, whenever subtractive and additive ideas do arise, potential biases in choice become another source of variance.

We believe this general framing circumvents the finer-grained distinctions that follow below about the exact role of expectations versus accessibility in experiments 1—5. We appreciate the rigorous exchange on this issue and believe that it has led to both a more general and less controversial account of the results.

That being said, I think there are a few ways in which I think this paper still falls short of its potential – and that can fortunately be rather easily remedied:

The conceptualization (in both the paper and the revision notes) of problem solving as a 2-stage process consisting of a consideration stage followed by an evaluation stage, with preference (and possibly expectations) only affecting the latter stage is both overly simplistic and logically inconsistent with other parts of the paper (see the first point I listed above).

Take the sentence (p. 5): “Preference-based explanations are relevant only when people consider both additive and subtractive actions.” That is simply not true. If additive strategies are more desirable (e.g., more rewarding) or more normative (e.g., what’s generally expected in these situations), then that will affect the generation of ideas, not just their evaluation. I urge the authors to remove this sentence and any other suggestion that preference or expectations only affect the evaluation stage.

We did not mean to suggest that preference or expectations only affect the evaluation stage. The text has been modified accordingly.

While we recognize that the reviewer’s point is not referring to this sentence alone, please note that we have revised to clarify that the key distinction is between what people think of and what they choose – for example, we have removed the phrase ‘preference-based explanation’ and refer in appropriate places to “choice.”

(line 64) “Low rates of subtraction raise two broad possibilities about the thought processes that produce them: people might generate both kinds of ideas and then disproportionately choose the additive ones; or they might overlook subtractive ideas altogether. While both phenomena may be contributing to the observed behavioral outcomes, we focused our subsequent research on potential differences at the idea-generation phase because it necessarily precedes explicit choice.”

In my previous review, I pointed out that the results of experiments 1-4 could be driven by participants’ expectations of what they are supposed to do. Some participants may implicitly expect that they’re supposed to add something, so those may only consider subtractive solutions when the instructions explicitly tell them it’s ok to do so. The authors now counter this with two arguments, neither of which is convincing to me. First, they state that the instructions made clear they could change it in any way (“alter it any way you want”). That of course doesn’t preclude a segment of participants to still assume that they’re supposed to add something. Second, they present a new study (S11) in which participants are explicitly presented with both subtractive and additive solutions and both types of solutions are judged as legitimate by participants, irrespective of instructions. However, explicitly presenting a

subtractive solution by itself signals to participants that this is an acceptable way to approach the problem, so this doesn't address the issue at all. This goes back to my previous point: the authors seem to assume that expectations can only affect evaluation, not idea generation. That is obviously not correct: If I implicitly assume that the experimenter is looking for additive solutions, then I will only generate additive solutions unless it is made explicit that subtractive ones should also be considered (either through instructions or by showing me subtractive solutions). I would encourage the authors to downplay study 1-4, to acknowledge that participants' expectations may have influenced the results, and not to present S11 as a rebuttal of the role of expectations (given that showing the subtractive solutions by itself changes expectations).

Thank you. We have clarified the contribution of study S11. It shows (descriptively) that, based on the instructions given, most people in experiment 1 likely would have recognized subtraction as a legitimate solution if they considered it. Inferentially, it fails to support a choice account. (We are careful not to say that it rules out a choice account.) To the Reviewer's point, though, we do not attempt to position it as informing the role of expectations (only choice).

I think the new study 5 is a good addition to the paper (as I noted earlier), but even though it is much subtler than studies 1-4, it is still possible that this manipulation increases subtractive solutions because it changes expectations. People who implicitly assume that they are supposed to offer additive solutions may change this assumption when they notice, in the practice trials, that each of the problem sets can be solved by changing a single constellation of squares. That is subtle, and unlikely to be a demand effect, but it may still be a way to challenge expectations. To be clear, I still think this is an interesting result. Moreover, participants' natural assumption that they are supposed to add rather than subtract is related to the central thesis of this paper. However, rather than positioning this study as showing that people only consider subtractive solutions when they dig deeper, I would recommend that the authors acknowledge that it may also lead them to question assumptions.

Thank you for the original feedback that prompted us to include this experimental design in the package. Our revised description says (line 170), "The repetition merely gave them more opportunities to recognize shortcomings of an additive approach." We think this accommodates the possibility of challenged expectations through experience and clearly illustrates how the finding advances the central thesis of a heuristic additive search strategy.

In the end, I'm wondering if "chronic accessibility" is the correct term for what the authors are proposing as the mechanism driving the reduced likelihood to generate subtractive solutions. Reduced accessibility suggests that it is harder to come up with (i.e., it is harder to access in memory). Indeed, the authors now state, as one of the possible drivers of this reduced accessibility, that "additive transformations may be incrementally easier to process." However, I don't think that's generally true. As I mentioned in my previous review, we do know that it is easier to generate counterfactuals that undo an action rather than add one. It seems plausible to me that it is easier to simply remove something that is already there than to come up with something new to be added. Moreover, I don't think that the "accessibility" terminology is essential for the authors' account. The abstract talks about "heuristic search strategies" and I think that is a more accurate (less speculative) way to describe it: people somehow default to additive solutions. These solutions have a privileged status and are more

likely to be spontaneously considered. It is about WHICH solutions you try to access rather than how EASY it is to access them.

Again, we agree that a focus on chronic accessibility may be distracting to general readers and possibly more controversial to specialists. As described in the revised project summary, what our results show is that people systematically limit their mental searches to additive transformations and consequently overlook subtractive transformations. We agree that it is useful to leave it at that level. (We also acknowledge that all of the “cognitive, cultural, and socioecological reasons” offered for why additive search might have become a heuristic in the first place (lines 71—86) are circumstantial given that none of the work cited focused on additive versus subtractive changes—but the hypothesis should hold, whatever the distal causes are.) We have used similar descriptive phrasing throughout the main text to avoid introducing inessential terms.

Thank you once again for your careful attention to our work. Your time and thought has made the empirical package stronger and the account clearer.

Referee #3 (Remarks to the Author):

Thank you for your conscientious revision – the improved theoretical grounding, clarified conceptualization, added experiments, and more direct discussion of implications have addressed my concerns and enhanced the potential contribution of this work. Nice work!

Thank you for your invaluable time and thought. We particularly appreciate your encouragement to more directly discuss implications. Doing so has strengthened the manuscript (and our thinking.)