The Recency Bias in the Attribution of Credit and Blame for Joint Work Outcomes

Andrzej Baranski[†] Manuel Muñoz-Herrera[‡]

April 25, 2024

Abstract

In production processes within modern firms and organizations, it is often the case that individuals contribute sequentially to achieve a common goal. Does the order of contributions affect attribution of credit or blame for the outcome of a joint production process? We design an experiment in which teams of three subjects collaborate sequentially on a building task that must be completed within a given time limit. Uninvolved reviewers evaluate the builders' performance after watching a video of the task, and decide which builder to hire. We find robust evidence of a recency bias for blame attribution: The hiring rates of the final builders are substantially lower when the team fails. However, there is no difference in hiring rates between first and last builders for successful teams. We control for perceived task difficulty, objective and subjective individual contribution perceptions, attention, and recall, none of which explain or dampen the order effect. Our findings have implications for teamwork, organizational design, and management practices of team performance evaluation.

JEL Classification: C90, J70, L23, M50, M19

Keywords: credit attribution, attribution bias, team production, joint production, laboratory experiment

We would like to thank Laila Al-Eisawi, Marielle Caballero and Amr Yakout for their excellent research assistance. Florian Engl, Urs Fischbacher, Alex Gotthard-Real, Moritz Janas and Robert Stuber offered very valuable suggestions. The authors are grateful for financial support from the Center for Behavioral Institutional Design and Tamkeen under the NYUAD Research Institute award for Project CG005.

[†]Division of Social Science and Center for Behavioral Institutional Design, New York University Abu Dhabi. *Email*: a.baranski@nyu.edu

[‡]Luxembourg Institute of Socio-Economic Research (LISER), Esch-sur-Alzetter, Luxembourg. *Email*: manuel.munoz@liser.lu

1 Introduction

In many production processes between partners, employees, or team members, contributions often occur sequentially. Various divisions within a firm are typically involved in the creation of a product or provision of a service. For shareholders and management, it is crucial to determine to what extent each contributing individual is responsible for the collective outcome of the joint production process (Holmstrom, 1982; Milgrom and Roberts, 1992; Winter, 2006; Lazear and Shaw, 2007). *Proper* attribution of responsibility to individuals collaborating in a task is especially important within firms, as promotions and bonuses are likely to be awarded to employees whose contributions are most recognized, while those blamed for negative outcomes can be penalized, demoted, or fired (Lazear and Gibbs, 2014; Villeval, 2022).[1] As such, failing to attribute credit for joint work may lead to feelings of inequity (Adams, 1963; Carrell and Dittrich, 1978), which can be deleterious to team or firm performance (Bradler et al., 2016).[2] Importantly, the problem of attribution of responsibility is also an issue of wide concern in political settings (Rudolph, 2003), tort and criminal law (Fincham and Jaspars, 1980; Robbennolt and Hans, 2016), and sports, where the outcomes are typically an aggregation of sequential, individual efforts.

In this paper, we study whether the order in which individuals contribute to a sequential production task impacts how their contributions are perceived and recognized by an evaluator (e.g., a manager). Are those who contribute early in the production process deemed to be *more causal* (i.e. primacy bias) or is the responsibility for the outcome attributed to last movers (i.e. recency bias)? Furthermore, does success or failure alter who is credited or blamed for an outcome?

Although the questions we pose in this study appear simple at first glance, answering them empirically, based on observational data, proves elusive. Most joint production settings that one may envision are beset with a series of confounding factors that prevent a clear comparison between tasks within a team or firm. Typically, tasks performed by individuals may differ in complexity or relevance. Employees are likely to differ in their effort provision, the quality of their contributions, and their sta-

[1] Even in the absence of financial incentives, people typically care about being adequately recognized in relation to their peers. For example, Kosfeld and Neckermann (2011) and Bradler et al. (2016) find that managers giving a purely symbolic award (a thank-you card) to good workers, elicit more effort from recipients.

[2] See Dalton and Landry (2020) for how the recent performance of an NBA player can have an outsized effect on the likelihood of being traded compared to the effect of more informative historical data.

tus (i.e., seniority, power, etc.). Importantly, when there are strong synergies or complementarities in production as happens in most firms, early contributors to the task can affect the overall productivity and ability to contribute of those participating later in the sequence. Thus, it becomes unclear how to objectively evaluate contributions to the task. Furthermore, evaluators' judgements can be *motivated*, leading to favoritism or discrimination in their assessments, or memory shortcomings can affect assessments due to imperfect recall.

In this article, we design a series of experiments (divided into two studies) to control for the aforementioned factors and isolate the temporal order of contributions. In our first study, uninvolved evaluators (hereafter referred to as *managers*) are shown a video of a team of three other subjects who are building a toy model sequentially (hereafter referred to as *builders*). In the video that managers watch, each builder puts together a predefined number of pieces exogenously assigned by the experimenter, and the next builder continues the construction from where the previous builder finished until the task is completed by the third builder. In order to collect a measure of individual performance, each builder was asked to build the entire toy model individually and separately. Managers were not informed of the individual performance.

We then formed teams ex post, by creating videos that showed three distinct builders completing the task in a sequence. Managers were shown a video in one of two possible conditions. In the success condition, the team of builders completed their task within a specified time limit. In the failure condition, they did not. Upon watching the video with the three builders in a sequence, managers were asked who they would hire to perform the same task for them, a response that we incentivized by compensating managers for selecting high-ability builders (i.e., those with fast individual completion time).

We find strong evidence of a *recency bias* in the attribution of blame for failed outcomes: the last builder is over 3 times less likely to be hired than the first builder (14% vs. 47%). There are no significant differences in hiring rates between first and last builders for successful teams (36% vs 40%), which highlights that the recency bias is outcome-dependent in our setting. We are able to cleanly rule out multiple confounding factors such as objective and perceived contribution, task difficulty, other-regarding preferences, and limits in memory and recall ability from managers. Our findings are further substantiated by the responses of subjects when asked who was responsible for the group outcome, where a similar pattern arises to that of our incentivized hiring decision.

To address important aspects that are central to managerial settings and that can affect judgments of responsibility attribution, we conducted two variants of our experiment in the first study. First, we considered a variation in which managers were paid a bonus contingent on the outcome of the team they evaluated (hereafter, *Involved* treatment). This is motivated by the observation that manager's payoffs may be contingent on the outcome of their team, which in turn can trigger an emotional response that exacerbates biases. Second, managers' decisions typically affect workers, which may trigger more thoughtful or careful considerations if managers have equity concerns (Adams, 1963). To this end, we designed a treatment in which hired builders were rewarded monetarily (hereafter, *Rewards* treatment). Consistent with our main results, we find a recency bias in blame attribution for failed outcomes and no significant order effect for successful outcomes, but there are qualitative differences in line with our expectations. The last builders in the failed teams are marginally less likely to be hired in the *Involved* treatment and marginally more likely in the *Rewards* treatment. These results highlight that causal attribution judgments can be affected by experienced disappointment when the outcome has material consequences over the evaluator and that equity concerns may mildly temper the recency bias.

A crucial aspect of team production that is missing intentionally by design in the experiments of our first study is that members of a team are likely to know about the performance of those preceding them in the production sequence. In our first study, team production videos were formed after builders finished the entire task individually. It is plausible that attributions of responsibility vary if evaluators believe that last movers in failed teams should have reacted by speeding up. For example, a third-party evaluator may argue that a last mover could have *saved the day* by exerting more effort to build faster, and hence consider the last mover's actions more causal when the team fails. Under this line of reasoning, a first mover would be spared the blame as he would have no way of reacting. Alternatively, in the case of success, a final mover may be perceived as overwhelmingly responsible. Hence, to increase the realism of our experiment, we conducted a second study in which builders know how their predecessors performed.

We find a strikingly similar recency bias for failed teams, with no evidence that last movers are credited more than first movers under success. This means that builders' awareness of past performance does not lead evaluators to judge failed last movers more harshly or more favorably under success. Taken together, the results of both studies posit a disadvantage for last movers when teams fail. Our findings have important implications for management practices in the evaluation of outcomes for joint work, an aspect we discuss further in our concluding remarks.

The remainder of this article proceeds as follows. In Section 2, we discuss the existing related work. Next, in Section 3, we present the experimental design followed by the results of Study 1 in Section 4. The results from the *Involved* and *Rewards* treatments are presented in Section 4.2. Section 5, reports the results from our complementary Study 2. In Section 6, we report the a text analysis of the reasons managers gave for why the attributed credit/blame and why they hired a given participant. We discuss our results and conclude in Section 7.

2 Previous Work

In the field of cognitive psychology, there is a large literature studying how the order in which information is presented to people affects their judgements. In these studies, subjects express their opinion about a specific proposition after being presented with several pieces of information, which may arrive sequentially. We refer the reader to the meta-analysis by Hogarth and Einhorn (1992), which documents evidence both of primacy (over-reliance on initial evidence or information) when the information is *simple* and recency (over-reliance on late information) when information is complex. We consider our experimental building task to be very simple, which would trigger primacy according to Hogarth and Einhorn's conceptual framework, but this is not what we find.

Closer to our experiment, several recent studies have focused on order effects in determining *causal judgements*. As defined by Reuter et al. (2014), "[c]ausal selection is the cognitive process through which one or more elements in a complex causal structure are singled out as actual causes of a certain effect" (p. 1). Spellman (1997) argues that causality is typically attributed to the event which raises the probability of occurrence of an outcome the most, given what has happened earlier. Vignette experiments by Henne et al. (2021) reveal that events that change the probability of the outcome the most, are deemed more causal. In the context of a collective decision-making problem, Bartling et al. (2015) find that pivotality affects responsibility attribution in sequential voting, with pivotal voters being blamed more for an unpopular outcome than non-pivotal voters by recipients of a collective decision. Anselm et al. (2022) find that pivotal voters are rewarded more often for supporting a fair outcome. Supporters after the pivotal voter are blamed and rewarded less, thus there is no recency or primacy. Engl (2022) developes a theory of causal responsibility attribution in two-stage games that helps explain the patters of behavior in Bartling et al. (2015) and Anselm et al. (2022).

There is a growing interest in understanding credit attribution for joint work in teams (Lazear and Shaw, 2007), especially in academic research collaborations (see Shen and Barabási (2014) for a review). Survey evidence suggests there are self-serving biases in credit claim (Herz et al., 2020), when there is no objective way of establishing percentage shares of contribution. Kornhaber et al. (2015) offers a review of articles discussing the challenges of co-authorship in the biomedical sciences. Importantly, none of these studies focus on manipulating the temporal order of contributions to a team or firm as we do here, and instead, point to the challenge of attributing individual credit for joint work.

Finally, experiments by Cappelen et al. (2017) and Almås et al. (2020) are examples of studies employing uninvolved parties that are informed about the outcomes of an individual production task. Workers in these experiments participate in real-effort tasks and the value of their production is partially determined by luck. Third parties make distributional decisions that affect the worker's payoffs. Besides the motivating question, there are several methodological differences in our settings. First, observers' decisions do not affect the workers' payoffs in our main experiments. We intentionally do so to turn off the effect that fairness concerns regarding payoff distribution may have. Second, observers witness the production process visually (i.e., watch a video) as opposed to being merely informed about it.[3]

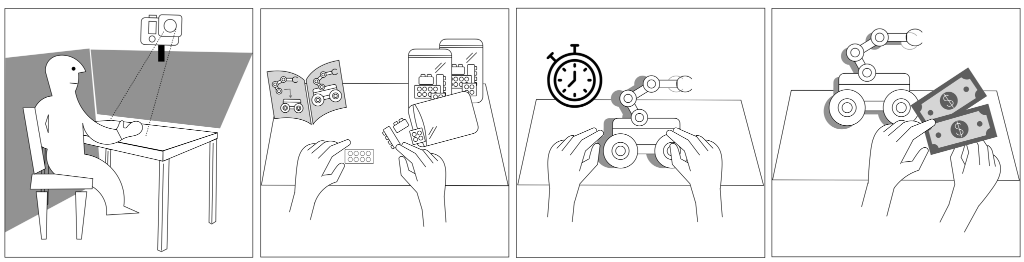
3 Experimental Design

The experiment consisted of two parts: a real-effort building task and an assessment of performance by evaluators, which leads to hiring decisions.

[3]See Khubulashvili et al. (2021) for a discussion on the effects of observing versus imagining someone's contribution.

Part 1: Building task

Builders were undergraduate students invited to the Social Science Experimental Laboratory at NYU Abu Dhabi to assemble a Lego® set. The sequence of events taking place in a session is illustrated in Figure 1. Participants come to the lab alone (individual sessions) and sit in a cubicle. A camera pointing to their hands is used to record videos of the building process (see Fig. 1a). The pieces of the Lego set are divided into three bags and each participant builds the set in a three-part sequence (see Fig. 1b): 40% of the pieces in bag 1, 20% in bag 2 and the remaining 40% in bag 3.[4] Once the set is completely built, the timer is stopped (see Fig. 1c). In addition to the show-up fee, a bonus was paid to builders who completed the task in less than 10 minutes.



(a) (b) (c) (d)

Figure 1: **Building Task (Part 1).** (a) Individual participants sit in a cubicle, where a camera records their hands while building a Lego set following a manual. (b) The Lego pieces are divided into three packs, one for each part of the sequence that the participant builds. (c) The timer is stopped when the set is completed. (d) If the task is completed in less than 10 minutes, the participant earns a bonus.

Interim: Forming the Teams and Creating Videos

Although builders completed the set individually, we presented managers in Part 2 with a video that resembled a sequential team production line. For this, we combined the videos of the three parts, each part completed by a different builder as shown in Figure 5. In the video, it is clearly stated that a new builder is starting a new part of the Lego set. At the end of every video showing a building sequence, evaluators are informed whether the team succeeded or failed to complete the building task on time

[4] One may argue that first and last builders faced different difficulty levels because first builders had more pieces at their disposal from where to search and then assemble. We further ensured that this was not the case by separating the pieces in bags for each part of the set. We also elicit managers' perceptions to control in our analysis for perceived difficulty (see instructions in Appendix A).

(see Fig. 2d).[5]

The goal of our videos was to show a realistic and easily comprehensible task for evaluators to assess. To reduce distraction and increase the quality of our responses, we shortened the videos so that only the moments when participants were putting pieces together were visible. We also sought to minimize differences between builders so that it would be difficult to identify *the true ability* of each builder. Thus, the duration of each part is not informative of the actual performance of each builder.[6]

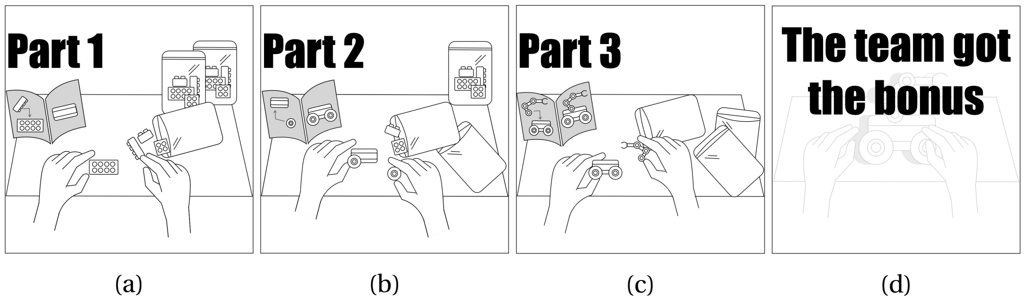


Figure 2: **Building sequence shown in the videos that managers evaluate.** Each manager observes the builders putting together the Lego set. When a builder starts a new part, this is clearly shown on the screen. At the end of the video, the managers are informed if the team failed to build the set in time to earn the bonus or succeeded in obtaining the bonus.

Part 2: Hiring

We invited an online sample of third-party evaluators (which we refer to as *managers*) who were paid a fixed fee of $2 to watch a video clip (as described above) and answer a few questions about the content of the video.[7] The task was described to managers as an evaluation of a team that had built a Lego set together and whose success depended on how long it took them to finish the task.[8] Half of the managers in each treatment were randomly assigned to watch a video in which the team failed, the other half watched a successful team.

[5] In total, we created 4 videos, two for each condition: success or failure. In our analysis, we control for which video was shown to the managers.

[6] Web links to the videos used can be found in the Online Appendix D. The approach in these videos ensures that differences in responsibility attribution and hiring, between success and failure, are a consequence of the order. However, this decision could impose some limitations in terms of external validity, and therefore, we complement our design in Study 2, with a new set of features.

[7] Subjects who did not correctly answer comprehension questions could not proceed with the experiment. In total, we recruited 1600 managers via Prolific, 400 per treatment.

[8] For the edited videos, we said a team succeeded if the total time (after speeding up the individual videos) is below three minutes. This is explained to the managers and displayed on the screen.

Manager's main task, after watching the video, was to choose which builder they would *hire*. Importantly, before making their decision, managers were informed that each builder had assembled the entire set individually and that some had succeeded and others failed to complete it in time (as explained in Section 3). Managers received a bonus payment of $2 if they hired a builder who successfully built the set on time in the individual session.

Managers were also informed that teams were formed ex post. This is a central feature of our design as it guarantees transparency, and prevents managers from assuming that final builders knew how much time was left before the time was out. Note that this aspect can substantially affect judgments since managers may believe that the final workers could increase their speed to beat the time limit when a team was lagging behind. As such, it would not be unreasonable to perceive a failed team as providing a signal that the final worker was *of low ability*. This would create a difference between successful and failed teams that could skew manager's choices in not hiring last movers under failure.

We also asked managers which builder they considered to be the most responsible for the team's outcome. Our conjecture was that the answers to this question and the hiring decision would be highly positively (negatively) correlated in success (failure) conditions.[9]

To control for the perception of contributions to the building task, we asked managers to estimate the number of pieces each builder had placed. Note that they were not informed about the share of pieces put together in each part of the video. Those managers who were within 5 Lego pieces of the correct amount would earn an additional bonus payment of $1. We also asked them which part of the assembly process they found was more difficult to build (unincentivized). Both of these measures allow us to investigate the determinants of hiring decisions. The accuracy of the answers on the assembled pieces also serves as a measure of attention and recall, because we can test if managers remember first and last movers' contributed pieces differently.

Finally, we ask managers to provide a verbal explanation of their reasons for hiring the builder they chose. Similarly, we asked them to explain the reasons behind the attribution of credit (in success) or blame (in failure) to the builder they chose to make responsible for the outcome. We coded these

[9] Part of our methodological inquiry was to evaluate if incentivizing the causal attribution questions could affect subjects' responses systematically compared to stated judgments. We find a high correlation and no systematic differences: causal attribution is positively correlated with hiring in success conditions and negatively in failure conditions.

responses to further evaluate the hiring decisions (see Section 6).

4 Results

4.1 Hiring Decisions

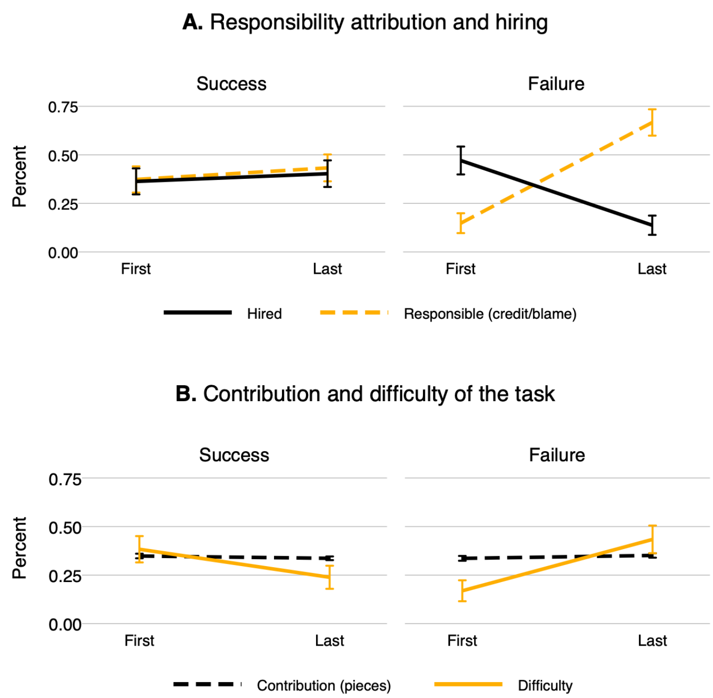


Figure 3: **Main Experimental Outcomes.** *Panel A:* The solid line (in both panels) displays the fraction of managers who hired the worker in either the First or Last position. The dashed line displays the fraction of managers who made the worker in the first or last position responsible for the success (left panel) or failure (right panel) of the outcome. *Panel B:* The dashed line displays the perceived contribution by the manager of each player to the task, expressed as the share of total pieces put together. The solid line displays the perceived difficulty by the managers of the task each player was given according to their position.

Figure 3.A shows the proportion of times that builders in each position (first or last) are hired by the manager, both when teams were successful (left) or failed (right). When teams succeed, first builders get hired 36% of the time while last builders 40% of the time (p=0.412, Chi-square test). When teams fail, the difference becomes significant: first builders get hired 47% of the time while last builders 14% of the time (p<0.001, Chi-square test). These findings provide direct evidence for the existence of a

recency bias in blame attribution when hiring: Last movers are disproportionately penalized for failure and not evidently prized for success.[10] The differential effect of the outcome on hiring probability of each builder is notable: failure leads to a larger hiring gap between the first and third builder ($p<0.001$).[11]

Repeating the same analysis as before, we find that managers are less likely to blame the first builder for the failure of a team than they are to blame the last mover (15% vs 67%, $p<0.001$, Chi-square test). However, as illustrated in Figure 3.A, managers appear to be equally likely to attribute credit between the first and last movers when teams succeed (37% vs 43%, $p=0.222$, Chi-square test), as illustrated in Figure 3.A. Moreover, when looking at the relation between hiring and attributing responsibility (credit or blame), we find a strong correlation between the hiring decision and credit (in success conditions) and no-hiring and blame (in failure conditions). The correlation coefficients are 0.500 ($p<0.001$) and -0.419 ($p<0.001$), respectively.[12]

Result 1. *There is a recency bias in the attribution of responsibility for failed teams but not for successful ones. The last builder in the sequence of a failed team is blamed more often and hired less often than the first builder, but there is no difference between first and last movers of successful teams.*

4.1.1 Controlling for perceived contributions and difficulty

In our experiment, our aim has been to control for each builder's *objective* contribution, as well as to maintain the task difficulty homogeneous between first and last builders. We now ask if managers effectively believe that contributions were equal between them, and if perceptions of difficulty vary within building positions or outcomes (success vs. failure). We also investigate whether managers' perceptions affect their hiring decisions and if they may explain the previously documented hiring bias.

With respect to our first question, it is natural to conjecture that if last builders are blamed more for

[10] The true performance of each builder (whether he builds the entire Lego set on time or not) is not significantly correlated with hiring decisions. The correlation coefficient is 0.018 ($p=0.535$).

[11] P-value obtained from a linear regression on the probability of hiring the last builder.

[12] This suggests that methodologically the incentivized measures, such as the hiring decision developed here, provide similar insights regarding credit attribution compared to stated judgements in non-incentivized questions. Thus, researchers seeking to study this phenomena in other settings where incentivizing responses can prove difficult may rely confidently on stated attributions of responsibility.

failures, it might be because managers also underestimate their material contribution to the building task. But this is not the case. There is no difference in the proportion of the pieces that managers believe first and last movers assembled (26% vs 25% for success, p=0.970; 24% vs 25% for failure, p=0.480).

We asked managers which builder had the most difficult part (position), with the option of stating that all were equally difficult. Figure 3.B shows that when teams are successful the first part is deemed the most difficult 38% of the time and the third part 24%, and the difference is significant (p=0.002). When teams are unsuccessful, the first part is perceived as the most difficult 17% of the time and the third part 43% (p<0.001). Across outcomes, 28% of managers state that all three parts were equally difficult.

	Success		**Failure**	
	I	**II**	**III**	**IV**
Builder 3	0.040	0.079	−0.333***	−0.350***
	(0.062)	(0.059)	(0.052)	(0.053)
Contribution		1.400***		0.851**
		(0.274)		(0.284)
Difficulty		0.117*		0.008
		(0.059)		(0.052)
Constant	0.363***	−0.173	0.476***	0.194
	(0.037)	(0.94)	(0.041)	(0.102)
# Obs.	402	400	378	376
# Groups	201	200	189	199
Video fixed effects	Yes	Yes	Yes	Yes

Table 1: **Linear Regressions for Hiring.** Linear regressions with standard errors clustered on individual managers (in parentheses). The dependent variable is probability of being hired by the manager in columns **I** and **II** for successful outcomes and in columns **III** and **IV** for failed outcomes. To test for recency/primacy biases, we focus only on managers hiring either the first or the last builder. ***, ** and * indicate statistical significance at the 0.001, 0.01 and 0.05 levels.

How do managers' perceptions of pieces contributed and task difficulty correlate with their hiring decisions and attribution of responsibility? Also, once we account for these perceptions, does the recency bias remain? Table 1 presents the results from linear regressions for each outcome variable in both success and failure conditions. Our results clearly indicate that perceived contributions are a key driver of hiring decisions and that perceived difficulty has no impact. Importantly, the recency bias in failed outcomes remains.

The results presented in this subsection help us rule out that managers' poor or selective memory can be driving our results. We also corroborate that subjects are trying to make objective assessments in hiring the highest contributor, and not just guessing. Furthermore, we checked whether the accuracy of subjects in assessing the total number of pieces that builders *contributed* differed by condition, but this is not the case. The Euclidean distance between the actual and reported values is 22.5 pieces on average and is not significantly different between success or failure conditions (p=0.522, obtained from a linear regression). In Appendix B, we replicate the analysis on hiring decisions, splitting managers into those whose accuracy on the actual contributions of the builders (number of pieces) was below and those above the median. Our findings remain for both groups of managers, which we interpret as our results not being driven by task attention.

Result 2. *On average, managers accurately perceive that the first and last builder contribute equally to the task, meaning there is no recency bias in assessing material contributions to the task. While contribution perceptions correlate positively with hiring decisions, these do not explain the recency bias in hiring.*

4.2 Follow-up Treatments

In this section, we present the results of additional treatments aimed at capturing two central aspects of managerial decision making: that the outcome of the team can affect managers' payoffs, and that their evaluations can affect workers' earnings.

In the *Involved* treatment, managers' payoffs varied depending on the outcome of the team they evaluated. Every aspect of the experiment remained the same, except that managers were endowed with $1 at the beginning of the experiment. They were informed that if the team they were assigned to evaluate was successful, they would earn an additional $1. Otherwise, they would pay a cost of $1. Because failed teams entail a monetary loss for managers, one reasonable conjecture is that this can trigger a stronger attribution of blame.

Figure 4 shows the hiring rates in all treatments, including the *Baseline* condition discussed in the previous section. The data show that managers in treatment *Involved* display a recency bias in blame attribution when teams fail (as in the *Baseline*). Importantly, managers are less likely to hire the last builder in treatment *Involved* relative to baseline, although the difference is not significant in success

(3.69%, p = 0.434, Chi-square test) or failure (4%, p = 0.217, Chi-square test).

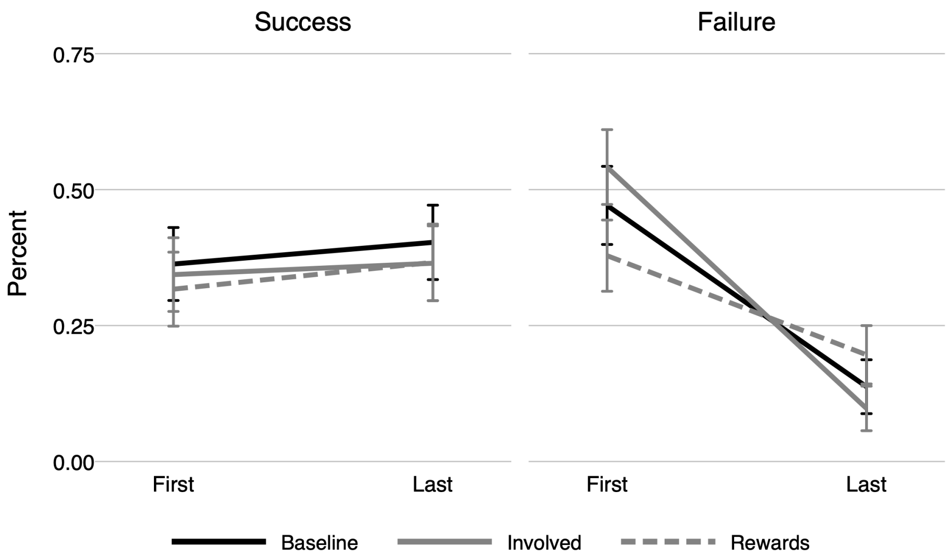


Figure 4: **Hiring decisions in Additional Treatments.** Each line displays the fraction of managers who hired the worker in either the first or last position in treatments Baseline (dark solid line), Involved (light solid line) and Rewards (dashed line). Hiring decisions are displayed separately for successful outcomes (left) and failed outcomes (right).

We also run a treatment in which managers' hiring decisions could carry rewards for workers. Managers were told that there was a 10% probability that we (the experimenters) would pay $1 to the worker they hired. We argue that this variation activates the possibility of manifesting and acting upon equity concerns to properly reward those deemed more responsible for a team's success and punish those deemed more responsible for a team's failure.

In the REWARDS treatment, managers are more likely to hire the last builder relative to the baseline in failure, although the difference is not significant in failure (5.87%, p = 0.116, Chi-square test) or success (3.69%, p = 0.458, Chi-square test).

Result 3. *The recency bias in hiring, in which last builders are hired less often than first builders when teams fail, persists even when payoffs of the managers are affected by the team's outcome or when managers' decisions carry rewards for workers.*

5 Study 2: Enhancing Realism in the Team Building Process

In the description of Study 1, we explained why we designed a stylized environment that controlled for most confounding factors that could influence the hiring decision, beyond the position in the building sequence. Study 1 can be conceived as a *wind-tunnel* where many ecological validity can be at stake. Naturally, this poses a trade-off between realism and experimental control. We conducted a second study with two objectives in mind. First, we aimed to increase the external validity of our experiment by bringing it closer to the *real-world* settings we have discussed to motivate our research questions. To this end, we varied the way in which the team production process occurred, as explained in the following. A second objective of Study 2 is to substantially increase our sample, both of building teams and managers.

As before, *builders* are invited to the lab to assemble a Lego set. But, unlike the first set of studies, three participants come to the lab at the same time and sit in separate cubicles (see Figure 5). First, to obtain a measure of individual performance, each builder completes a Lego set individually (we did not record videos, only completion times). Individual participants were offered a bonus for their individual task if their time was among the fastest 50% of the builders.

Once all three participants finish building their individual set, they are told they have to build a new set, but this time as a team. Builders are randomly assigned a position in the building sequence.

As in study 1, we had a dedicated cubicle with a camera to record the building process. When builders completed their assigned part, the next builder would come to the dedicated cubicle.

We recorded the sequential building task with the same specifications as in Study 1 (showing only the hands of the builders). The videos were accelerated, and we made clear that there were three parts with three distinct builders. However, unlike videos in Study 1, no other edits were made. This means that every action by the builders can be observed by the managers, for example, pauses or mistakes. Hence, subjects in the role of managers are potentially able to perceive the ability of the builders.[13]

The builders in the sequential building task were also incentivized to complete the set as fast as possible. They were offered a bonus if their teams was among the fastest 25% or 50% of the teams.

[13]Recall that, in Study 1, the videos were edited to show only moments were bricks were being assembled. We cut out moments when builders were searching for a piece to place or staring at the building manual.

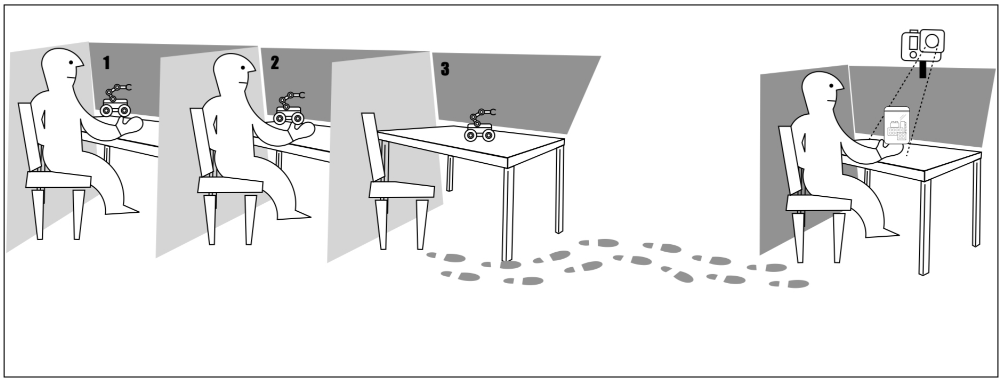


Figure 5: **Team building task.** Three participants come to the lab at the same time and sit in separate cubicles. Each of them builds individually a Lego set first (without video recording). Then, they are randomly assigned a position to build a Lego set together in sequence. Each of them goes individually to a separate station where a camera records their hands building their corresponding part of a different Lego set. The set pieces are divided in three packs that the participants build, one each. The timer is stopped when the set is completed by the participant in the final position. If the task is finished below a time threshold (either in the 50% or 25% fastest among all groups) the participants in the team earn a bonus.

They were told that the exact threshold would be randomly selected at the end of the entire data collection process (once all teams have participated). This design feature allowed us to produce videos for some teams under both conditions (success and failure) without incurring in deception, which further strengthens our identification of a possible bias mediated by the outcome while holding all else constant. For example, a team ranked in the 70^{th} percentile, could be shown as a failed team or a successful one.

Managers participated in assessing only one video in one condition (success or failure) as in Study 1. They earned a bonus if they hired a builder who was among the fastest in the individual task. Managers did not observe or have any information about the performance in the individual task. All they could see was the video of the team-building task.

A secondary objective of this study was to investigate the robustness of the results presented in Study 1, which were based on the assessment of 4 different teams (2 videos for each condition). In Study 2, we have 6 different teams. In total, we collected 621 responses from the evaluators using the same pool as in Study 1. No subject participated twice as per their unique Prolific IDs.

5.1 Results Study 2

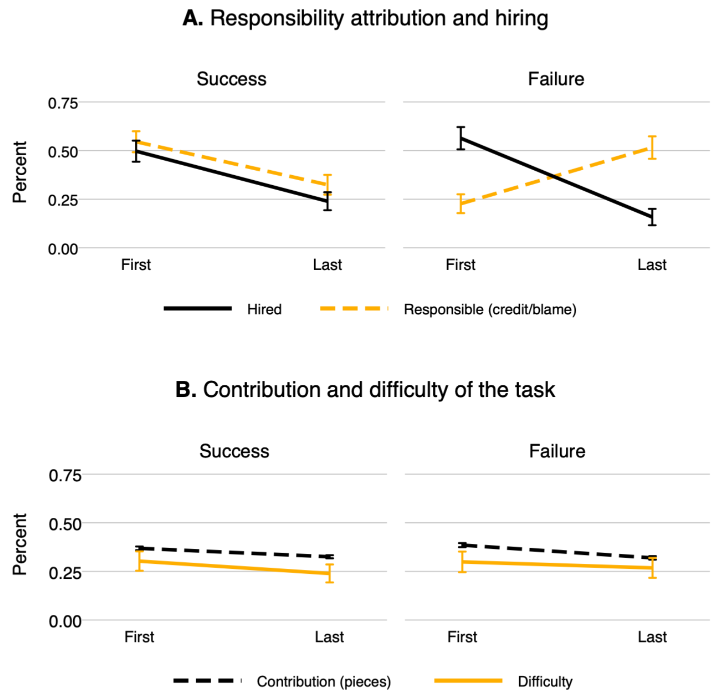


Figure 6: **Experimental Outcomes.** *Panel A:* The solid line (in both panels) displays the fraction of managers who hired the worker in either the First or Last position. The dashed line displays the fraction of managers who made the worker in the first or last position responsible for the success (left panel) or failure (right panel) of the outcome. *Panel B:* The dashed line displays the perceived contribution by the manager of each player to the task, expressed as the share of total pieces put together. The solid line displays the perceived difficulty by the managers of the task each player was given according to their position.

Figure 6 depicts the hiring rates of the first and last builders, for each condition. Conducting the same regression analysis as for Study 1 (see Table 2), we find that the third builder is significantly less likely to be hired overall, both in success and failure.

Importantly, when we control for the number of Lego pieces placed per minute (our objective measure of building ability), the recency bias remains for failed teams but disappears for successful ones, in line with the findings of Study 1 (see columns III and VI in Table 2).

	Success			Failure		
	I	II	III	IV	V	VI
Builder 3	−0.258***	−0.213***	−0.205	−0.405***	−0.376***	−0.612**
	(0.045)	(0.055)	(0.155)	(0.044)	(0.056)	(0.186)
Contribution		1.204***	1.202***		0.777**	0.812**
		(0.246)	(0.253)		(0.277)	(0.280)
Difficulty		0.108*	0.108*		−0.121**	−0.107*
		(0.047)	(0.047)		(0.042)	(0.042)
Constant	0.487***	−0.034	−0.038	0.514***	0.232	0.344*
	(0.041)	(0.163)	(0.171)	(0.046)	(0.119)	(0.149)
# Obs.	660	660	660	582	582	582
# Groups	330	330	330	291	291	291
Video fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Objective ability	No	No	Yes	No	No	Yes

Table 2: **Linear Regressions for Hiring.** Linear regressions with standard errors clustered on individual managers (in parentheses). The dependent variable is probability of being hired by the manager in columns **I** and **II** for successful outcomes and in columns **III** and **IV** for failed outcomes. To test for recency/primacy biases, we focus only on managers hiring either the first or the last builder. ***, ** and * indicate statistical significance at the 0.001, 0.01 and 0.05 levels.

6 Stated Motives for Hiring Decisions

After the hiring decisions were made, subjects were asked at the end of the experiment to explain the reasons for their choice. Our goal is to identify if the motives that subjects state can illuminate the reason why the first builder in the sequence is hired more often than the last one under failure.

We hired three independent coders to classify managers' responses using seven binary categories. These categories are whether the builder was hired because of the perceived importance of his/her part, perceived difficulty, because of the builder's speed, the amount of pieces contributed, how focused the builder appeared to be, or whether the decision was random (see details in Appendix E).[14] In Figure 7 we report the distribution of motives for successful (panel A) and failed (panel B) building outcomes. For the motives question we had a response rate of 82% in success and 81% in failure.

The reasons managers provide for their hiring decisions are strikingly similar regardless of whether they hire the first or the third mover, or whether the outcome was successful or a failure. Managers say that they chose who they perceived as the fastest 50% of the time: 51% for both first and third movers in successful outcomes (p 0.90), and 60% and 47% with failed outcomes (p 0.008). This is followed by

[14] With three coders, discrepancies are broken by majority rule.

focus as the second most frequent motive, which is stated 19% in success and 23% in failure.

We coded one additional category concerning a possible reason why the last builders are not hired at the same rate as first builders under failure, which has to do with our motivation for Study 2. It might be that managers believe that if a team is progressing slowly it is the responsibility of the last builder to pick up the pace. Note that this believe should not arise in study 1 as teams were formed expost and the third builder had no way of knowing the performance of the preceding builders. Thus, the responsibility motive should only arise in study 2 where builders are aware of the time elapsed at any point during the building task. We find that this reason is not a relevant motive in neither of the studies, as it is mentioned only 0.6% of the time (i.e., 7 participants across studies).

In a nutshell, the stated responses for hiring motives do not provide a clear reason for the lower hiring rate of the third builder under failure. Instead, we view these responses as strengthening the notion that a bias is in place: Despite providing similar motives for hiring, hiring rates differ for third and first builders under failure.

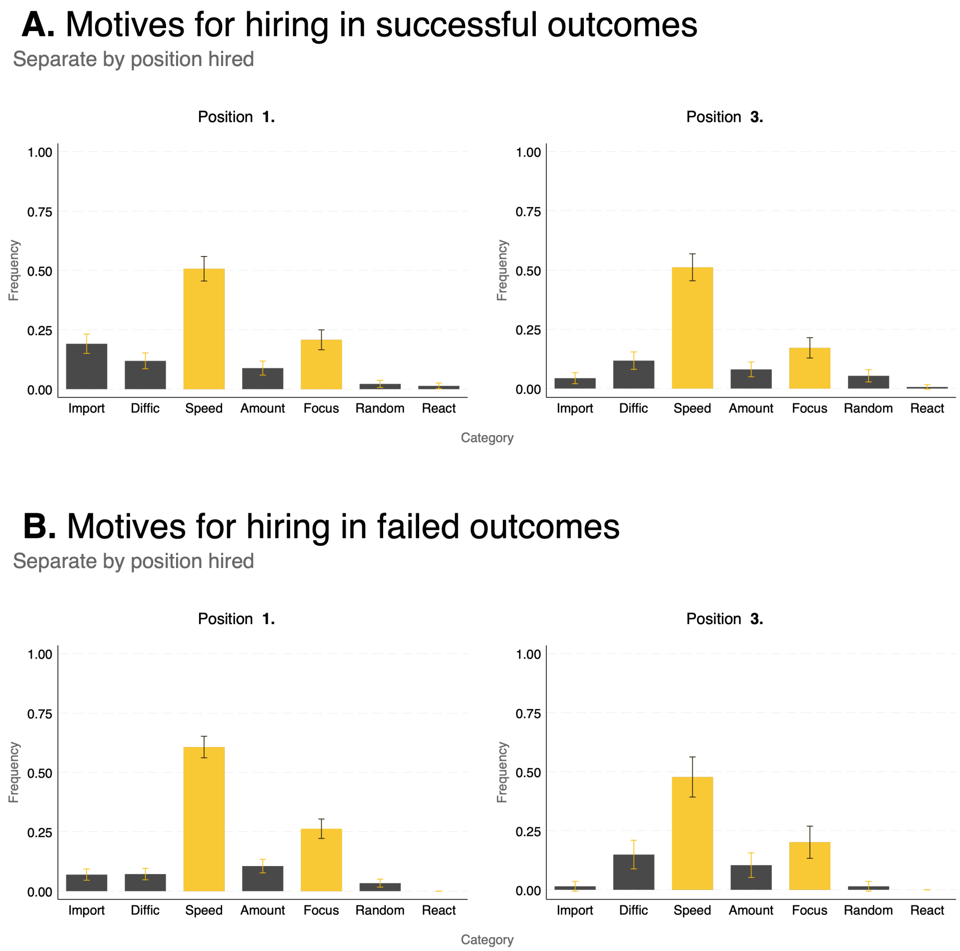


Figure 7: **Motives for hiring.** The figure reports the average classification of the reasons managers gave for why they hired a builder after a successful outcome (Panel A) and after a failed outcome (Panel B).

7 Discussion and Concluding Remarks

In this study, we present evidence of how the order of contributions to a team task affects blame attribution for the outcome. The last contributor to a sequential task is judged to be responsible more often when the team being evaluated has failed to achieve its objective. The judgements of attribution of responsibility by third parties correlate strongly with their hiring decision, a choice that carries material consequences in our study. Thus, we not only uncover a bias in judgement, but find that it can have deleterious material consequences.

Importantly, we rule out several possible mechanisms that could potentially explain the bias we identify in our data. First, the bias does not arise due to imperfect recall or impaired memory because we do not find differences in the perception of pieces contributed between the first and last mover. Second, the bias does not operate through selective recall dependent on the outcome because evaluators' subjective assessments of contributions do not differ between success and failure conditions. Third, we control for task difficulty by ensuring that the first and third parts are equally difficult, and find no perceived differences in this dimension by managers. Fourth, we rule out the possibility that the last builders are deemed to be more responsible for failure because they are expected to *save the team*. By design in Study 1 we rule this cannot be a motive because builders have no way of reacting to a slow team. And importantly, in Study 2 where this concern may arise, hiring behavior looks virtually identical to that of Study 1. Finally, hiring decisions are not affected by reputation effects because there are no repeated interactions, nor do they have an effect on the effort of builders, so that hiring decisions have no strategic role to play in fostering efficiency.

In addition to the behavioral patterns we uncover, our contribution to the existing literature is also methodological. We designed a task that objectively controls material contributions to the team task and keeps difficulty relatively unchanged between builders. In our first study, early contributors cannot directly affect the productivity of late contributors: a quick builder in the beginning is as conducive to team success as a quick builder at the end. These features allow us to isolate the effect of the order of contributions quite transparently. Importantly, the hiring decision avoids any confounding factors that may arise due to social preferences for fairness or concerns for efficiency. This is because managers' decisions do not affect the builders in our main treatment.

The strength of the recency bias in the attribution of failure is strong and stable in subsequent treatments that mimic important conditions in managerial settings. We explored two conditions that enhance the external validity of our findings: When managers' payoffs depend on the outcome of the team they evaluate and when managers' hiring decisions carry monetary rewards for workers. In both cases, the recency bias persists. In addition, there is an indication that last builders are even more likely to be blamed for failed outcomes when the manager's payoffs are affected. One possible explanation is that disappointment may trigger an emotional response that heightens the recency bias.

In the treatment in which hiring decisions carry potential rewards for builders, we find a moderately lower blame on last builders of failed teams. One potential reason for the tempering of the bias is that evaluators may be concerned with fairness in terms of equity, that is, rewarding justly for the work done. Because subjects report no differences in pieces assembled between first and last builders, they are likely to view the workers as equally deserving of the monetary reward and this mildly dampens the impulse to not hire the last builder in failed teams.

Our findings may have important implications for management practices. Many business endeavors typically start with an idea generation phase that takes place at high levels within a firm's hierarchy. If high-ranking executives are deemed the initiators of projects but other employees down the hierarchy contribute at later stages in their execution, it may happen that failed endeavors are less likely to be blamed on the initiators. As such, if a rencecy bias for failed outcomes is at play, it can shield management and high-ranking executives from blame, which can lead to firing or sanctioning the *wrong* culprits.

In general, those in charge of evaluating employee performance in team-based firms should account for a potential bias in the attribution of responsibility, which is expected to be more pronounced in low performance teams. Not being able to identify those responsible for a team's failure can lead to an unfair and inefficient allocation of resources within organizations.

Several questions remain to be answered, which may enhance our understanding of the recency bias that we have uncovered. Can workers anticipate this pattern of judgment by managers? And if given a choice, which place in a production sequence would workers or teammates place themselves? A second area to explore concerns complementarities in production, which are central to teams and firms. Note that in our task, the first builders' efforts have no direct impact on the productivity of the subsequent builders. It remains to be studied whether causal judgments can be affected by the presence of synergies in joint production or not. We leave these and other important questions for future research.

References

Adams, J. S. (1963). Towards an understanding of inequity. *The Journal of Abnormal and Social Psychology*, 67(5):422.

Almås, I., Cappelen, A. W., and Tungodden, B. (2020). Cutthroat capitalism versus cuddly socialism: Are americans more meritocratic and efficiency-seeking than scandinavians? *Journal of Political Economy*, 128(5):1753–1788.

Anselm, R., Bhatia, D., Fischbacher, U., and Hausfeld, J. (2022). Blame and praise: Responsibility attribution patterns in decision chains. Thurgau Institute of Economics and Department of Economics at University of Konstanz Research Paper Series available at: https://www.twi-kreuzlingen.ch/research/workingpaper/. Last accessed October 9, 2022.

Bartling, B., Fischbacher, U., and Schudy, S. (2015). Pivotality and responsibility attribution in sequential voting. *Journal of Public Economics*, 128:133–139.

Bradler, C., Dur, R., Neckermann, S., and Non, A. (2016). Employee recognition and performance: A field experiment. *Management Science*, 62(11):3085–3099.

Cappelen, A. W., Moene, K. O., Skjelbred, S.-E., and Tungodden, B. (2017). The merit primacy effect. NHH Dept. of Economics Discussion Paper available at SSRN: https://ssrn.com/abstract=2963504. Last accessed 19 June 2022.

Carrell, M. R. and Dittrich, J. E. (1978). Equity theory: The recent literature, methodological considerations, and new directions. *Academy of management review*, 3(2):202–210.

Dalton, M. and Landry, P. (2020). 'overattention'to first-hand experience in hiring decisions: Evidence from professional basketball. *Journal of Economic Behavior & Organization*, 175:98–113.

Engl, F. (2022). A theory of causal responsibility attribution. *Available at SSRN 2932769*.

Fincham, F. D. and Jaspars, J. M. (1980). Attribution of responsibility: From man the scientist to man as lawyer. *Advances in experimental social psychology*, 13:81–138.

Henne, P., Kulesza, A., Perez, K., and Houcek, A. (2021). Counterfactual thinking and recency effects in causal judgment. *Cognition*, 212:104708.

Herz, N., Dan, O., Censor, N., and Bar-Haim, Y. (2020). Authors overestimate their contribution to scientific work, demonstrating a strong bias. *Proceedings of the National Academy of Sciences*, 117(12):6282–6285.

Hogarth, R. M. and Einhorn, H. J. (1992). Order effects in belief updating: The belief-adjustment model. *Cognitive psychology*, 24(1):1–55.

Holmstrom, B. (1982). Moral hazard in teams. *The Bell journal of economics*, pages 324–340.

Khubulashvili, R., Linardi, S., and Wang, X. (2021). The role of effort cost perception in outcome bias. *Working paper*.

Kornhaber, R. A., McLean, L. M., and Baber, R. J. (2015). Ongoing ethical issues concerning authorship in biomedical journals: an integrative review. *International journal of nanomedicine*, 10:4837.

Kosfeld, M. and Neckermann, S. (2011). Getting more work for nothing? symbolic awards and worker performance. *American Economic Journal: Microeconomics*, 3(3):86–99.

Lazear, E. P. and Gibbs, M. (2014). *Personnel economics in practice*. John Wiley & Sons.

Lazear, E. P. and Shaw, K. L. (2007). Personnel economics: The economist's view of human resources. *Journal of economic perspectives*, 21(4):91–114.

Milgrom, P. and Roberts, J. (1992). *Economics, organization and management*. Prentice-Hall Englewood Cliffs, NJ.

Reuter, K., Kirfel, L., Van Riel, R., and Barlassina, L. (2014). The good, the bad, and the timely: how temporal order and moral judgment influence causal selection. *Frontiers in psychology*, 5:1336.

Robbennolt, J. K. and Hans, V. P. (2016). *The psychology of tort law*. Springer.

Rudolph, T. J. (2003). Who's responsible for the economy? the formation and consequences of responsibility attributions. *American Journal of Political Science*, 47(4):698–713.

Shen, H.-W. and Barabási, A.-L. (2014). Collective credit allocation in science. *Proceedings of the National Academy of Sciences*, 111(34):12325–12330.

Spellman, B. A. (1997). Crediting causality. *Journal of Experimental Psychology: General*, 126(4):323.

Villeval, M. C. (2022). The cognitive and motivational effects of performance feedback. In Erikson, T., editor, *Encyclopedia of Labor Studies*, page Edward Elgar Publishers.

Winter, E. (2006). Optimal incentives for sequential production processes. *The RAND Journal of Economics*, 37(2):376–390.

Online Appendix:

The Recency Bias in the Attribution of Responsibility for Joint Work

A Instructions

Below we present the instructions for the experiment. We will indicate when a question applies to a specific treatment. For example, **[Success]** when the manager observes a team who has successfully completed the task, and **[Failure]** in the other case. Also, we will indicate if the answer options for a question were displayed in random order by using *[*randomized order]*.

WELCOME

You are participating in a study on economic decision-making. Typically, the study takes about **8 minutes** to complete.

For completing the study, you will receive $2.00. In addition, you will be able to earn **bonus payments** of up to $3.00 more. You will be paid only if you complete the entire study.

The study is anonymous. Hence, your identity will not be revealed to others and the identity of others will not be revealed to you.

Next, you will see the instructions. **Please read the instructions carefully as they describe how your earnings are determined.**

The main task is to carefully watch a short video. You will be asked questions to confirm that you have watch it with attention. **If you answer these questions incorrectly, you will be excluded from the study and you won't be eligible for payment.**

By continuing to the next screen, you consent to participate in this study. For more details about your consent, click on "See consent form".

- See consent form

[page break]

We asked participants in a different study to build parts of a Lego set. We then grouped these participants into teams, and gave a bonus to those teams for which the total time of building the Lego set was at most 3 minutes.

The following video shows you a team of participants building the Lego set. Please watch the video carefully, as we will ask you a few attention questions about the task the participants are performing.

[page break]

Please watch carefully the following video.

Note: Please pay close attention to the task participants are performing. Skipping this stage or leaving before the video ends will make you ineligible for payment.

*[*The video is displayed in this question]*

[page break]

Did the participants in the team earn the bonus by successfully building the Lego set in time?

- Yes
- No

[page break]

How many participants composed the team (built the lego set) in the video you watched? (*Hint: the number of participants is equal to the number of parts in the video*)

[options from 0 to 10]

[page break]

Please indicate how many pieces you think **each participant** put together. If your answer is within 5 units of the correct amount of pieces for each participant, you will receive a bonus of **1 USD**.

- Participant 1
- Participant 2
- Participant 3

[page break]

Please indicate **how difficult** you think the different parts of building the Lego set were *[*randomized order]*

- All three parts were equally difficult
- The first part was the most difficult
- The second part was the most difficult
- The third part was the most difficult

[page break]

[**Success**]

Which of the three participants in the team was **the most responsible for building** the lego set in less than three minutes *[*randomized order]*

- Participant 1
- Participant 2
- Participant 3

[page break]

Please tell us why did you say *[Chosen participant]* was the most responsible for building the Lego set in time to earn the bonus

[page break]

[**Failure**]

Which of the three participants in the team was **the most responsible for not building** the lego set in less than three minutes *[*randomized order]*

- Participant 1
- Participant 2
- Participant 3

[page break]

Please tell us why did you say *[Chosen participant]* was the most responsible for not building the Lego set in time to earn the bonus

[page break]

In the video you watched, each participant built one of the three parts of the Lego set. We also asked each participant to build the other two parts, meaning that each of them built the entire Lego set by him or herself.

You have the opportunity to **hire one of the three participants** in the video. If the person you hire has built the entire Lego set (the three parts) in time for him/her to earn the bonus, you will earn an additional **bonus of 2 USD**. Please indicate which of the three participant you choose to hire.

- Participant 1
- Participant 2
- Participant 3

[page break]

Please tell us why did you hire *[Hired participant]*

[page break]

Please indicate what was the gender of each of the participants

- Participant 1: Male / Female
- Participant 2: Male / Female
- Participant 3: Male / Female

[page break]

Before concluding, we would like you to answer a final set of questions.

What is your gender?

- Male
- Female
- Other

[page break]

What is your race / ethnicity? Select all that apply.

- White
- Black
- Latino
- Asian
- Native American
- Other

[page break]

How old are you (in years)?

[page break]

What is your highest educational degree?

- Less than high school
- High school graduate
- 2 year college/university degree

- 4 year college/university degree
- Masters degree or equivalent
- Doctorate or equivalent

[page break]

Are you currently employed?

- Yes
- No

[page break]

Are you currently a student?

- Yes
- No

[page break]

In which state do you currently reside?

- Alabama
- ...
- I do not reside in the United States

[page break]

Did you vote in the last election?

- Yes
- No

[page break]

Generally speaking, do you usually think of yourself as a:

- Republican
- Democrat
- Libertarian
- Independent
- Other

[page break]

Here is a 7-point scale on which political views that people might hold are arranged from extremely liberal (left) to extremely conservative (right). Where would you place yourself on this scale?

[page break]

You have reached the end of the study.

Please do not discuss the procedures or content of this study with other participants.

Your bonus payment

In the next 48 hours we will review your responses to the questions that had a bonus payment, and determine whether you earn the **bonus payment**. To receive your payment you must submit your prolific ID on the next page.

B Baseline treatment supporting analysis

We replicate the analysis from hiring for responsibility attribution. Table B1 reports linear probability models testing the decisions to attribute responsibility over the success or failure of the group's outcome. In all regressions the independent variables are dummies for the final builders, so that the first builder is the omitted category. Moreover, we use individual-level random effects in all regressions to cluster decisions at the manager's level. The dependent variable is the probability of assigning responsibility to a builder, for the case of Success in columns **I-II** and Failure in columns **III-IV**.

	Success		**Failure**	
	I	**II**	**III**	**IV**
Builder 3	0.060	0.106	0.519***	0.514***
	(0.063)	(0.058)	(0.054)	(0.056)
Contribution		1.520***		−0.091
		(0.275)		(0.283)
Difficulty		0.226***		0.053
		(0.059)		(0.060)
Constant	0.373***	−0.242*	0.148***	0.165
	(0.034)	(0.95)	(0.026)	(0.098)
# Obs.	402	400	378	376
# Groups	201	200	189	199

Table B1: **Linear Regressions for Responsibility attribution.** Linear regressions with standard errors clustered on individual managers (in parenthesis). The dependent variable is probability of assigning responsibility of the group's outcome by the manager in columns **I** and **II** for successful outcomes and in columns **III** and **IV** for failed outcomes. ***, ** and * indicate statistical significance at the 0.001, 0.01 and 0.05 levels.

In Table B2, we replicate the main analysis on hiring decisions. Managers are split into those whose accuracy on the actual contributions of the builders (number of pieces) was below the median in columns **I-II, V-VI** and above the median in columns **III-IV, VII-VIII**.

	Success				Failure			
	Below median		*Above Median*		*Below median*		*Above Median*	
	I	**II**	**III**	**IV**	**V**	**VI**	**VII**	**VIII**
Builder 3	0.050	0.096	0.030	0.054	−0.263***	−0.299***	−0.404***	−0.430***
	(0.091)	(0.081)	(0.085)	(0.085)	(0.074)	(0.074)	(0.072)	(0.071)
Contribution		1.709***		0.560		1.309***		−0.352
		(0.320)		(0.681)		(0.352)		(0.565)
Difficulty		0.113		0.140		−0.051		0.098
		(0.079)		(0.088)		(0.072)		(0.072)
Constant	0.380***	−0.264*	0.347***	0.098	0.421***	0.013	0.521***	0.626**
	(0.052)	(0.198)	(0.051)	(0.121)	(0.049)	(0.106)	(0.048)	(0.233)
# Obs.	200	198	202	202	190	188	188	188
# Groups	100	99	101	101	95	94	94	94

Table B2: **Linear Regressions for hiring - Median split on accuracy of contributions.** Linear regressions with standard errors clustered on individual managers (in parenthesis). The dependent variable is probability of being hired by the manager in columns **I-IV** for successful outcomes and in columns **V-VIII** for failed outcomes. Managers are split into those whose accuracy on the actual contributions of the builders (number of pieces) was below the median in columns **I-II, V-VI** and above the median in columns **III-IV, VII-VIII**. ***, ** and * indicate statistical significance at the 0.001, 0.01 and 0.05 levels.

C Additional treatments

C.1 Regressions

We replicate the analysis for hiring decisions as presented for the Baseline in Table 1. We report linear probability models testing hiring decisions for treatment Involved in Table C1 and for treatment Rewards in Table C2. In all regressions the independent variables are dummies for the final builders, so that the first builder is the omitted category. Moreover, we use individual-level random effects in all regressions to cluster decisions at the manager's level. The dependent variable is the probability of hiring a builder, for the case of Success in columns **I-II** and Failure in columns **III-IV**.

	Success		**Failure**	
	I	**II**	**III**	**IV**
Builder 3	0.021	0.003	−0.444***	−0.496***
	(0.061)	(0.058)	(0.047)	(0.049)
Contribution		0.876*		0.370
		(0.434)		(0.234)
Difficulty		0.269***		0.094*
		(0.064)		(0.045)
Constant	0.344***	−0.020	0.541***	0.402***
	(0.034)	(0.140)	(0.035)	(0.087)
# Obs.	384	382	410	410
# Groups	192	191	205	205

Table C1: **Linear Regressions for hiring in treatment INVOLVED.** Linear regressions with standard errors clustered on individual managers (in parenthesis). The dependent variable is probability of being hired by the manager in columns **I** and **II** for successful outcomes and in columns **III** and **IV** for failed outcomes. ***, ** and * indicate statistical significance at the 0.001, 0.01 and 0.05 levels.

Table C3 reports linear probability models testing hiring decisions by pooling treatments BASELINE, INVOLVED, and REWARDS together. In all regressions the independent variables are dummies for the final builders, so that the first builder is the omitted category. Moreover, we use individual-level random effects in all regressions to cluster decisions at the manager's level. The dependent variable is the probability of hiring a builder, for the case of Success in column **I** and Failure in column **II**.

	Success		Failure	
	I	**II**	**III**	**IV**
Builder 3	0.049	0.065	−0.182***	−0.206***
	(0.061)	(0.057)	(0.050)	(0.057)
Contribution		0.875**		0.244
		(0.313)		(0.435)
Difficulty		0.277***		0.038
		(0.066)		(0.060)
Constant	0.317***	−0.050	0.379***	0.294*
	(0.035)	(0.106)	(0.033)	(0.139)
# Obs.	366	366	428	428
# Groups	183	183	214	214

Table C2: **Linear Regressions for hiring in treatment REWARDS.** Linear regressions with standard errors clustered on individual managers (in parenthesis). The dependent variable is probability of being hired by the manager in columns **I** and **II** for successful outcomes and in columns **III** and **IV** for failed outcomes. ***, ** and * indicate statistical significance at the 0.001, 0.01 and 0.05 levels.

C.2 Figures

Below we include figures of the main choices and outcomes for treatments INVOLVED (Figure C1) and REWARDS (Figure C2).

	Success	**Failure**
	I	**II**
Builder 3	0.040	−0.333***
	(0.062)	(0.051)
Builder 1 X INVOLVED	−0.019	0.071
	(0.048)	(0.050)
Builder 1 X REWARDS	−0.046	−0.092
	(0.048)	(0.049)
Builder 3 X INVOLVED	−0.038	−0.040
	(0.049)	(0.033)
Builder 3 X REWARDS	−0.037	0.059
	(0.050)	(0.037)
Constant	0.363***	0.471***
	(0.034)	(0.036)
# Obs.	1152	1216
# Groups	576	608

Table C3: **Linear Regressions for hiring pooling all treatments.** Linear regressions with standard errors clustered on individual managers (in parenthesis). The dependent variable is probability of being hired by the manager in column **I** for successful outcomes and in column **II** for failed outcomes. ***, ** and * indicate statistical significance at the 0.001, 0.01 and 0.05 levels.

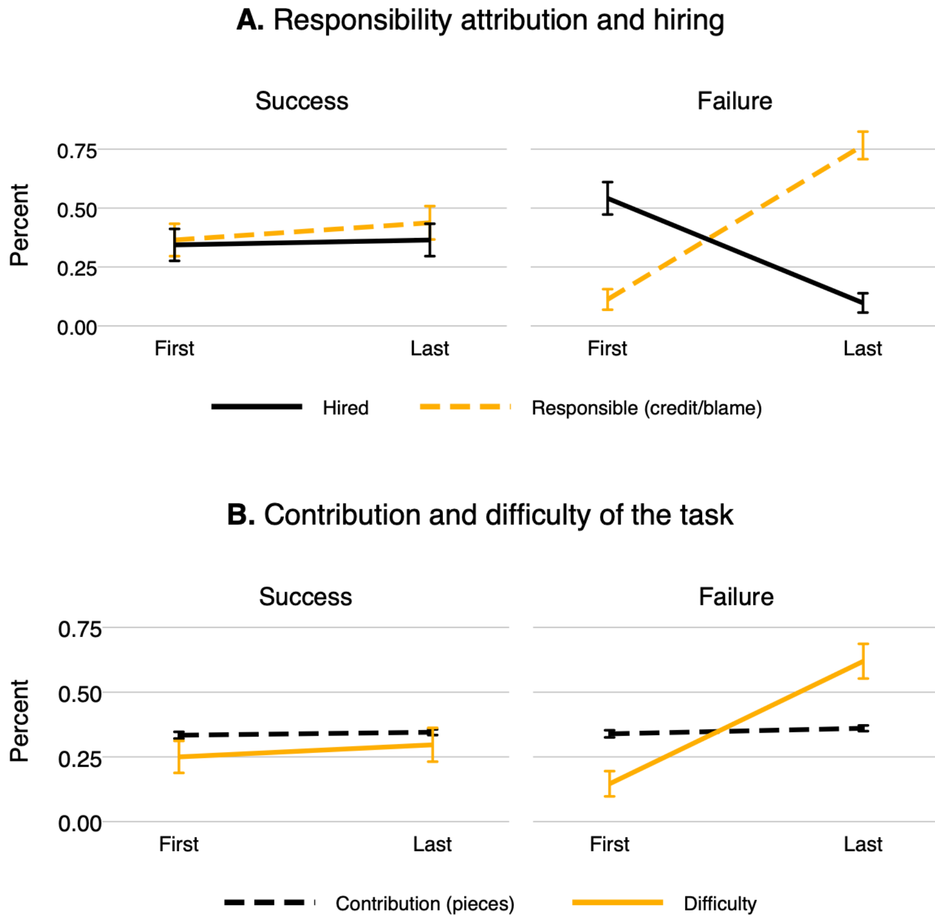


Figure C1: **Main Experimental Outcomes in treatment INVOLVED.** *Panel A:* The solid line (in both panels) displays the fraction of managers who hired the worker in either the First or Last position. The dashed line displays the fraction of managers who made the worker in the First or Last position responsible for the success (left panel) or failure (right panel) of the outcome. *Panel B:* The dashed line displays the perceived contribution by the manager of each player to the task, expressed as the share of total pieces put together. The solid line displays the perceived difficulty by the managers of the task each player was given according to their position.

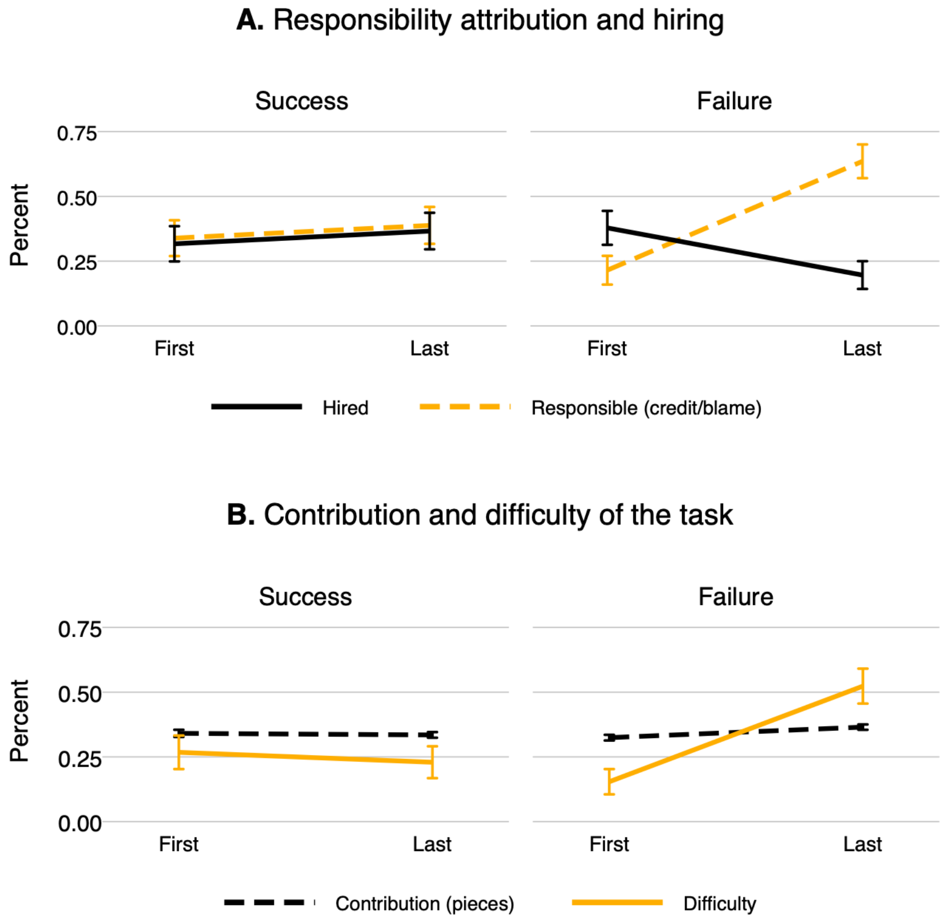


Figure C2: **Main Experimental Outcomes in treatment Rewards.** *Panel A:* The solid line (in both panels) displays the fraction of managers who hired the worker in either the First or Last position. The dashed line displays the fraction of managers who made the worker in the First or Last position responsible for the success (left panel) or failure (right panel) of the outcome. *Panel B:* The dashed line displays the perceived contribution by the manager of each player to the task, expressed as the share of total pieces put together. The solid line displays the perceived difficulty by the managers of the task each player was given according to their position.

D Links to videos

Below we include a link to the videos the evaluators watched in the main treatments:

- **Failure:**
  - https://www.youtube.com/watch?v=7R_DfCd9rUQ&t=2s
  - https://www.youtube.com/watch?v=mx_t_eK9yr0&t=1s
- **Success:**
  - https://www.youtube.com/watch?v=ZbALYGj4Vdc&t=3s
  - https://www.youtube.com/watch?v=jNNN9gKJ5G4

Below we include a link to the videos the evaluators watched in the treatments with real groups:

- **Failure:**
  - https://www.youtube.com/watch?v=XyzEDlx5PUA&t=1s
  - https://www.youtube.com/watch?v=7Ot2o687zNs
  - https://www.youtube.com/watch?v=ah5C1FlSOM8
  - https://www.youtube.com/watch?v=nOqfk1vKoHQ
  - https://www.youtube.com/watch?v=qy2BYFiWgZc
  - https://www.youtube.com/watch?v=wvSriQu32GA
- **Success:**
  - https://www.youtube.com/watch?v=r1LzY-4UQFk
  - https://www.youtube.com/watch?v=UkWnukazX24
  - https://www.youtube.com/watch?v=2MoPYoea7_w
  - https://www.youtube.com/watch?v=9ItjXh4SPXI
  - https://www.youtube.com/watch?v=wxcUOO_bNGM

- https://www.youtube.com/watch?v=7v3opnWjig4
- https://www.youtube.com/watch?v=j971ozWuj2c&t=1s

E Coding free form responses

Below we include the instructions we gave three independent coders on the categories to identify from the free-form responses managers when asked what motivated their choice.

Coding Task for Hiring Decision

An online sample of responders watched a video of a team of three players building a Lego set. The lego set was shown in the video as built in a sequence of three parts (one for each player).

Responders were asked which of the players they would like to hire and why.

You will be reading the responses to reasons for their hiring decisions. We would like to code the reasons participants give into 8 different categories.

Categories:

1. **Importance**: if the reason is that the subject completed the most important part. Keywords may be "crucial" "critical" "essential", "foundation" etc. It also includes statements such as "It was the last part", "Did the finishing touches", "Started it all" where it is presumed that this was an important part. Enter 1 if the text falls into this category, 0 otherwise.

2. **Difficulty**: if the reason is that the subject completed the most difficult part of the model. Keywords can include "challenging", "complicated". Enter 1 if the text falls into this category, 0 otherwise.

3. **Speed**: if the reason is that the subject was the fastest. Other keywords may include "efficient" "quick", etc. Enter 1 if the text falls into this category, 0 otherwise.

4. **Amount**: If the reason is that the subject placed the highest amount of pieces. Enter 1 if the text falls into this category, 0 otherwise.

5. **Focus**: If the reason is that the subject was very focused, did not get distracted, made no mistakes, was very skillful, had high ability. Enter 1 if the text falls into this category, 0 otherwise.

6. **Random**: when it is stated that the decision was made at random, with no particular motive. Also applies when responders state they "felt", had a "hunch", or similar phrasing. Enter 1 if the text falls into this category, 0 otherwise.
7. **Reaction**: the builder reacted by going faster given what she / he was handed. He/she sped up due to the slow progress of those preceding.

Note: a given text may fall into two categories or more, or may not fall into any category (all 0). That is perfectly fine, these are not mutually exclusive nor exhaustive.