Dynamics of Causal Attribution

Dana Kulzhabayeva (dana.kulzhabayeva@mail.utoronto.ca)
Department of Psychology, University of Toronto, 100 St. George Street
Toronto, ON M5S 3G3 CA

Joseph J. Williams (williams@cs.toronto.edu)
Department of Computer Science, Department of Psychology, University of Toronto,
40 St. George Street, Toronto, ON M5S 2E4 CA

David Danks (ddanks@ucsd.edu)
Halıcıoğlu Data Science Institute, Department of Philosophy, University of California San Diego
3234 Matthews Lane, La Jolla, CA 92093 US

Abstract

Attribution theory aims to explain people's judgments about the cause of some behavior or outcome, often involving other people. The theory has proven to be broadly applicable and points towards important aspects of human cognition. This relevance is perhaps unsurprising given that attribution theory is a type of causal inference. However, there has been relatively little work on attribution theory in relation to causal *learning*. More specifically, previous literature has mostly examined attributions and their behavioral and motivational outcomes following a single observation, rather than capturing the dynamics of causal attribution (i.e., how those judgments shift as people observe more vignettes and thereby learn about the situation). We thus ran an exploratory study using a vignette design to investigate whether attributions and their outcomes change across multiple instances of observation and behavior adaptation.

Keywords: attribution theory; learning dynamics; causal reasoning; causal learning

Introduction

Attribution theory has been a foundational theory in social psychology since (at least) Heider (1958). Over the span of five decades, over four million studies and articles have been published extending the theory, creating a space of what can now be called "attribution theories" that have been connected to motivation, behavior, affect, and social cognition.

The basic tenet of attribution theory is that we can deconstruct the causal structure of any outcome to three distinct dimensions: locus, controllability, and stability (Collins, Martin, Ashmore, & Ross, 1973; Weiner, 1979; Weiner, 1983). *Locus* is whether an outcome is due to internal factors (those within a person) or external factors (those within the environment). For example, if a person dreams of becoming a singer but does poorly on their first live performance, then the cause could be internal (singing ability, nerves) or external (microphone quality, energy of the audience). *Controllability* is whether the causal factors underlying an outcome are controllable (from the perspective of the interpreter) or uncontrollable; can I curb my nerves by my actions, or am I cursed with a debilitating anxiety disorder? Finally, *stability* captures the regularity and repeatability of the relevant factors. For example, the poor performance could be due to a freak malfunction with the AV equipment (unstable), or a lack of skill that will not change without intervention (stable).

At its core, attribution theory is about the idea that people seek out causal explanations for the outcomes they experience. This "drive to explain" serves two main purposes: 1) *learning*: it allows us to interpret outcomes and thereby better model the causal relationships and affordances in our physical and social environment. 2) *prediction*: it allows us to anticipate the effectiveness of our actions, and adjust our future beliefs, motivation, and behavior accordingly (Weiner, 1985). Both functions work to help us effectively manage ourselves and our environment in the short and long term.

These distinct functions have (arguably) been studied in two different spheres of research: *attributional* theories and *attribution* theories (Martinko & Thomson, 1998). Attributional theories provide frameworks for understanding the *processes* that individuals use to determine the cause of their own and others' behavior. For example, Kelley's covariation model states that people determine the cause of an individual's behavior based on three factors: consensus (whether others behave similarly in similar situations), distinctiveness (whether this person behaves differently in different situations), and consistency (whether this person always behaves similarly in this situation) (Kelley, 1973). Combinations of these factors determine whether a behavior is attributed to the person or to the situation. For example, if a new coworker cracks a joke on their first day and everyone laughs, you cannot conclude whether they happened to tell a funny joke, or they are an inherently funny person. However, if the coworker is funny across different situations and others find them funny, you can infer that this is a trait rather than situational circumstance. Kelley's model aligns with fundamental principles of causal learning, which suggest that individuals learn about causality by studying patterns of covariation in the environment, ideally building (over time) more accurate causal models and thus more accurate inferences (Kelley, 1973).

In contrast, attribution theories examine the behavioral, affective, and motivational *consequences* of attributions. A major figure in this area is Weiner, whose research supports

In L. K. Samuelson, S. L. Frank, M. Toneva, A. Mackey, & E. Hazeltine (Eds.), *Proceedings of the 46th Annual Conference of the Cognitive Science Society.*

the idea that the causal inferences people make for their outcomes have a significant impact on their subsequent motivation and achievement strivings (Weiner, 1985). Decades of experiments have since found that individuals who attribute an outcome to controllable factors are more likely to persist in their goals, increase their efforts, and demonstrate cognitive resourcefulness and flexibility compared to those who attribute an outcome to uncontrollable factors (Abramson, Seligman, & Teasdale, 1978; Frieze & Weiner, 1971; Weiner, Russell, & Lerman, 1978). Additionally, studies have found that the stability dimension influences changes in people's expectancy of success on a task following an outcome; individuals are likely to experience decreases in expectancy following failures and increases in expectancy following successes, but only if those outcomes were due to constant factors (like skill) rather than fluctuating factors (like mood) (Lewin et al., 1944).

Attribution theory makes strong predictions—that have largely been confirmed—about the relationship between people's causal interpretations of a situation and their subsequent motivation and behavior. Attribution theory has been applied to help understand and improve people's outcomes across numerous fields including education, organizational behavior, and clinical psychology (Brewin, 1985; Heider, 1958; Martinko, Harvey, & Dasborough, 2011; Peterson et al., 1982). However, the current attribution literature is limited by the fact that prior studies have almost exclusively focused on attributions and outcomes for a single instance of feedback. Classic attribution paradigms often involve participants completing tasks in which their success or failure is experimentally manipulated, followed by measures of participants' attributions, as well as additional behavioral, motivational, and affective measures (for example, see Weiner et al., 1971). Even studies on attributional theories, which examine the processes by which attributions are formed, largely present participants with all relevant pieces of information at once and ask them to form attributions, rather than measuring attributions as they arise and change in response to multiple vignettes (for example, see Orvis, Cunningham, & Kelley, 1975). This design risks studying something other than the attributional process that occurs in naturalistic settings.

There is significant potential value in measuring, modeling, and understanding the evolving and dynamic processes between attributions and their behavioral and motivational outcomes. In a naturalistic context, the attributional process involves a constant flow of feedback between a person's causal inferences, behaviors, and outcomes. For example, if a student fails a school quiz and attributes the result to their effort, then they will be more likely to study longer for the next quiz. This behavior will now influence the subsequent outcomes the student experiences, either in the predicted or unpredicted direction. A predicted outcome (i.e., they pass) confirms the student's understanding of the causal structure of the situation. However, an unpredicted outcome (i.e., another failure) would presumably prompt the student to act and gather new data, and potentially revise their causal inference. For example, a student might try alternative strategies if studying longer did not give the desired result. And if studying is repeatedly unsuccessful, then they may update their attribution to something other than effort as the cause. The dynamic relationship between attributions and their outcomes introduces complexity but also allows more richness in our understanding of how various factors influence the trajectory of people's causal learning and outcomes over time.

In this paper, we report an exploratory study that aims to bridge the gap between attribution and attributional theories by investigating how people's attributions for a negative achievement outcome may change across multiple instances of feedback.

Relevant Prior Literature

The current literature on attributional dynamics is sparse, including only a handful of older studies of change in attributional judgements over time. For example, Moore et al. (1979) examined whether people's attributions of their own behavior to dispositional or situational factors change over time. They had participants describe themselves across four behaviors (friendliness, talkativeness, assertiveness, and nervousness), after which half of the participants completed an attributional questionnaire rating their own behavior, while the other half did the questionnaire 3 weeks later. They found a main effect of time, such that people tended to make more dispositional and less situational attributions for their own behavior at time two compared to time one. In contrast, Wells (1982) had participants watch scenarios of an actor's behavior across situations, rather than their own. Some participants were asked to make attributions while others simply watched. All participants were asked to recall the actor's behavior one week later and those who made attributions had better recall.

A previously-noted (Lau, 1984) limitation of the above studies is that they seem to be capturing the effects of memory processes; the attributional changes may be solely a result of salience and information available due to recall. In contrast, Bernstein, Stephan, and Davis (1979) collected students' expectancies and attributions of their performance on three successive tests during a semester class. They found a significant change in expectancies from optimistically high to much lower across time. However, they did not correlate changes in students' expectancies with changes in attributions, instead analyzing each time point separately.

Finally, Lau (1984) investigated whether observers' attributions for success and failure change over repeated observations by analyzing sportswriters' attributions for outcomes of National Football League games across a season, specifically focusing on the stability and locus dimensions. The study examined changes in the hedonic bias effect, where individuals tend to attribute favorable outcomes to internal factors, and unfavorable outcomes to external factors. The study found a decrease in the hedonic bias effect in the locus dimension over time, such that wins were attributed less internally and losses less externally as the season progressed.

There was also an effect on the stability dimension, such that wins were judged as less stable than losses early in the season, but the pattern reversed later in the season. The study by Lau comes closest to addressing the question of how attributions change across multiple instances of observation. However, observing causal inference or causal learning in this context is difficult because explaining a sports team's loss or victory can be akin to explaining the fluctuations of the stock market. We can measure changes in attribution due to cognitive heuristics and biases, but this setting is too noisy to be able to examine people's causal learning and identify potential factors that affect attributional dynamics.

Past studies do not formulate their research questions in terms of causal learning, nor do they systematically manipulate factors that could lead to changes in causal inference, such as covariation between vignettes or explicit changes in behavioral strategies between observations. Finally, these studies do not systematically investigate how various behavioral and motivational measures change alongside changes in attributions.

These questions are crucial to explore because the current theoretical and experimental literature on attribution theory argues that there is a consistent relationship between a person's attributions and their behavior, affect, and motivation. This relationship is the basis for interventions and applications that use attribution theory to achieve real-world outcomes. Without an understanding of how attributions and outcomes interact over time, we might, e.g., give recommendations or design interventions that end up being harmful in the long run. For example, studies find that attributing outcomes to unstable causes leads to resilient expectancy and increased motivation following a single instance of feedback (Weiner, 1985). However, it is possible that repeated attributions to unstable causes may lead to maladaptive behavioral strategies. For example, a student who performs poorly once might feel better telling themselves that they got unlucky with the test questions, as this attribution can protect their expectancy and lead to perseverance. However, a student who repeatedly tells themselves they did poorly due to luck is also likely to ignore alternate strategies and not increase their efforts. In the long run, these attributions will lead to negative outcomes despite initially positive interpretations. A string of negative outcomes may lead to ever more harmful attributions (e.g., a student concluding that their repeated failure is due to internal, stable traits like intelligence). Moreover, between outcome and attribution, and attribution and behavior, there exists a vast space in which people's biases, motivated reasoning, and interpretations will be influenced by culture, demographics, and individual differences. Certainly, if the salient information is pointed out to us (as in most attributional studies), we can extrapolate from a single instance attribution to a trait-level cause. However, in real-world settings people must determine salience on their own. There are many factors outside the principles of covariance that could play a role.

In light of these gaps and lacunae, we designed an exploratory study using a vignette design to investigate whether people's attributions of a fictitious character's outcomes change across multiple instances of failure. We hypothesized that unstable outcomes would be judged as stable across multiple instances due to the repetition of failure. We also hypothesized that participants' attributions of controllability would shift from controllable to uncontrollable; as participants observe a person attempt to exert control over a situation and fail, they may be likely to make the inference that the cause is uncontrollable. We predict no change in the locus dimension across trials due to our experimental design. The locus dimension requires more complex inferences compared to the other two dimensions; for example, according to Kelley's covariation model, attribution of an outcome to a situation versus a disposition requires at least three pieces of information (consistency, distinctiveness, consensus). However, introducing this many variables would have led to a highly complicated design, therefore locus information was provided to participants directly without necessitating inference. We also measured participants' judgements of the character's subsequent behavior and motivation following each instance of failure. We predicted that participants' rating that the character will adjust their behavior would decrease over time, and that there would be a decrease across the motivation ratings.

Experiment

In the study, participants were shown a sequence of instances (i.e., vignettes) in which a fictitious character fails to achieve a goal. The cause of failure was manipulated between participants using the attributional dimensions: locus, controllability, and stability (2x2x2 factorial design). In each condition, participants saw eight subsequent instances from the character's progress on their goal across a four-week period. They were asked to judge the attributional structure of the outcomes, as well as the character's subsequent behaviors, motivation, and beliefs.

Participants

Ninety-six participants aged 18–45 were recruited for the experiment. Sixteen were excluded from analysis either because they did not complete all trials of the study, or a subset of their data was not recorded due to a technical error. All participants were recruited from Prolific and received financial compensation for their participation. Participants were screened based on their current country of residence (United States and Canada), and primary language (English).

Procedure

The study was conducted online using the Qualtrics survey platform. At the start of the study all participants were presented with a scenario about a fictitious character named "Nika", a university student who signs up to participate in a research study. The study is investigating marathon running and requires her to try to achieve a goal of running regularly for four weeks. The participants are presented with Nika's

responses to an entry survey in which she sets a concrete goal ("I will run on Tuesday and Thursdays for one hour at the park nearby, once I get home from university") and rates both her expectancy of completing the goal, and her sense of specific and general self-efficacy in relation to the goal (all rating are on a 1-7 Likert scale, Nika responds to all with a rating of 7).

Participants are then presented with the first instance: on Tuesday, Nika attempts to achieve her goal and fails. The cause for her failure is manipulated across the three attributional dimensions. Participants are asked to select their attributions for locus, stability, and controllability (ex. "Was the cause of Nika's failure Internal (something about Nika, or something Nika did) or external (something about someone else, or some external event"). Participants were then presented with three slider scales and asked to rate the probability out of 100 that Nika will: 1) quit, 2) try again using the same strategy (nothing changes), or 3) try again but change something about her strategy, skills, or environment. Next, participants were asked to rate the probability out of 100 that Nika would: 1) change her strategy, 2) change the social or physical environment, 3) practice/develop a skill or ability, 4) practice/develop a meta-skill, or 5) not change anything. Finally, participants were asked to judge how Nika's expectancy, specific self-efficacy, and general self-efficacy ratings might change in light of her failing to reach her goal.

Participants observed eight vignettes in total, with Nika failing in all of them. The cause for Nika's failure is kept consistent within-participant across all eight vignettes, though the wording changed depending on participants' responses to attributional measures on the previous vignette. Participants provided the same set of ratings after each vignette.

Attributional Manipulations

Eight conditions were designed for each of the eight possible causal structures (2(locus) x 2(controllability) x 2(stability) design). Each condition outlined a cause for Nika's failure to achieve her running goal. For example, one said: "on the first Tuesday, Nika failed to achieve her goal. She failed because she had multiple exams earlier that day, and when she got home, she felt unmotivated to do her scheduled run." This cause is internal (she felt unmotivated), controllable (Nika could have done her run if she wanted to), and unstable (Nika felt unmotivated because she happened to have a lot of exams that day, so this cause is unlikely to be constant).

The goal of including all eight possible conditions was to detect the average trend in stability and controllability changes across different kinds of causes. We assume the processes underlying causal learning and inference are not dependent on the specific cause of failure, but rather operate on more general principles. Namely, we hypothesized that after repeated failure, any unstable causes would be judged as stable because participants observe them to repeat. Additionally, any controllable causes would be judged as uncontrollable as participants observe Nika attempt to exert control and fail.

The eight conditions each contained eight vignettes, reflecting consecutive instances in which Nika failed. The cause of failure was identical within participants (i.e., within each condition), but the vignettes were structured so that participants' attributions for the previous outcome influenced the contents of the next vignette, in an effort to reflect the way causal learning might occur in a naturalistic context. For example, suppose a participant judged that "On the first Tuesday, Nika failed to do her planned run because she had pressing course assignments to finish" was due to an external, controllable, and unstable cause. To enable participants to test the stability of the situation with no other variation, the subsequent vignette would simply repeat the scenario: "On Thursday, Nika failed to do her planned run again, because she had pressing course assignments to finish." In general, we avoided introducing any behavioral changes on vignettes judged to be unstable.

In contrast, if participants judged a vignette to be controllable, then on the subsequent vignette they would see the same failure scenario but with a behavioral adjustment. For example, "Nika tried to practice time management by scheduling time for course work that did not coincide with her running time. However, Nika found herself falling behind schedule and still having pressing homework to complete. As a result, she failed to complete her planned run." In these vignettes, Nika would cycle between changes in strategy, skill development, and environment, so participants could see her try a variety of behavioral adaptations. Overall, participants would all be exposed to eight vignettes with the character failing the same cause, with minor variations in text.

Results

Attribution Measures. As this was an exploratory study with a limited sample size (n=80), power limitations restricted the analysis to examining changes in participants' responses between the first and eighth vignette for the attributional measures. There was a significant change in participants' attributions of locus, such that participants were equally likely to rate the cause of the character's failure as internal or external in the first vignette and more likely to rate it as internal in the last one (McNemar's test) χ^2 (1) = 30.25, $p < .001$. There was also a significant change in participants' controllability attributions, χ^2 (1) = 8.03, $p < .004$: they made equal controllable and uncontrollable attributions in vignette one and more controllable attributions in vignette eight. Finally, there was a significant change in the stability dimension, χ^2 (1) = 19.36, $p < .001$, as participants were more likely to rate the cause of the character's failure as stable in vignette eight than in vignette one.

Behavioral Measures. Similar to the attributional measures, analysis of the behavioral measures was restricted to examining differences between the first and eighth vignette. A series of repeated measures t-tests revealed significant

changes across all but one of the behavioral measures. Participants' probability rating that Nika would quit her goal significantly increased ($t(79) = -6.23$, $p < .001$), while their rating that Nika would try again but change something about her strategy, skills, or environment significantly decreased ($t(79) = 6.04$, $p < .001$). That is, participants were more likely to think Nika would try again and change her approach following her first failure, but more likely to believe she would quit following her eighth failure. There was no significant change in their rating that Nika would try again without changing anything.

There were significant decreases in participants' probability ratings that Nika would change strategy ($t(79) = 2.53$, $p = .013$), change her social or physical environment ($t(79) = 3.04$, $p = .003$), develop a skill or ability ($t(79) = 3.94$, $p < .001$), and develop a meta-skill ($t(79) = 5.11$, $p < .001$). This pattern suggests that participants believed Nika might try to make behavioral adjustments to succeed in her goal after her first failure, but believed that Nika would be unlikely to make any behavioral adjustments after eight failures. In fact, there was a significant increase in participants' rating that Nika would not change anything ($t(79) = -7.01$, $p < .001$). These findings can be interpreted in relation to the "quit" rating in the first scale: if participants believe that Nika is most likely to quit, then they are also likely to believe she will not adjust her behavioral strategy.

Finally, there were significant decreases in participants' judgements of Nika's expectancy ($t(79) = 12.89$, $p < .001$), specific self-efficacy, ($t(79) = 11.59$, $p < .001$), and general self-efficacy ($t(79) = 10.74$, $p < .001$) between the first and eighth vignette. As Nika continues to fail across the four weeks, participants come to believe that she is likely to lose faith in the achievability of her goal, and that her self-efficacy beliefs in this task specifically, as well as other tasks in general, is likely to go decrease.

Discussion

We were interested in measuring how participants' attributions for repeated failure outcomes, and subsequent behavioral changes, shifted across multiple instances of goal-relevant feedback. We found that participants' stability attributions changed in the expected direction, shifting from unstable to stable across the eight vignettes. This result makes intuitive sense, as the stability dimension is defined by whether a cause repeats or fluctuates. As Nika continues to reliably fail due to the same cause, it seems reasonable to infer that the cause is stable. There was an unpredicted shift in the locus dimension, however, as participants' attributions shifted from external to internal across the eight vignettes. One possibility is that this shift was due to the experimental design. In scenarios where the cause for failure is internal, locus might remain consistent throughout simply because it is difficult to frame an internal cause as external without additional information. For example, if the vignettes stated that Nika failed because of a lack of self-discipline, then it is difficult to identify an external cause (unless one is introduced). In contrast, external causes can be framed as internal, especially over time. For example, if Nika failed to run because she had to take care of her ill mother, then although the cause is initially external, over time the persistent nature of the failure can be attributed to any number of internal traits (e.g., lack of time management skills, insufficient motivation).

This finding highlights a limitation of the vignette-style experimental design often used in attribution studies. In such controlled paradigms, it is up to the experimenter to highlight the salient aspects of a situation for participants, thereby potentially leading them to a conclusion or specific attributional profile. For instance, a study condition that provides an internal cause typically removes the influence of external factors (which have been hidden or 'filtered out' for participants). However, in naturalistic settings, external factors are present, and could impact the direction in which participants' attributions change.

Our study also found a significant change in the controllability dimension, but it was not in the predicted direction. Our hypothesis was that as Nika demonstrated efforts to change her behavioral strategies across vignettes (thereby exerting control), participants would infer that her repeating failure must be due to uncontrollable factors. Instead, participants were more likely to attribute the cause of Nika's failure to *controllable* rather than uncontrollable factors over time. A possible interpretation is that our intuitive theories of cognition tend to cast internal factors as being inherently controllable. For example, factors such as one's motivation, emotional regulation, self-discipline, or time management skills are all considered to be controllable. Since we found participants tended to attribute outcomes to internal factors over time, a consequence may be that the outcomes became viewed as controllable as well.

This finding cuts against our initial prediction for the mechanism underlying this change as well; we expected that participants might judge both internal and external causes as uncontrollable over time because of a change in their *interpretation* of the same event. For example, a lack of motivation might be interpreted as a situational circumstance, but after multiple vignettes, a participant might infer the person has a deeper problem with motivation (e.g., perhaps they suffer from depression). The latter is generally considered uncontrollable and would align with our hypotheses for that dimension. However, participants minimized uncontrollable interpretations, choosing to direct locus to self-regulation skills that were within Nika's control. This could highlight the presence of a form of the fundamental attribution error in which participants overweighted the likelihood of controllable factors in another person's failure.

The second set of findings from our study pertained to the behavioral and motivational outcomes participants predicted for Nika. The data aligned with our hypotheses that participants ratings that Nika would exhibit behavioral changes would decrease with time, as well as their ratings of Nika's expectancy, specific self-efficacy, and general self-efficacy. At face value, these findings seem to align with

prior studies which find that attributions of failure to internal and stable factors are likely to lead to reductions in effort and quitting (Seligman, Abramson, Semmel, & von Baeyer, 1979; Weiner, 1985). However, upon closer examination, the interpretation of these findings in relation to past work is less straightforward because previous studies tend to implicitly extrapolate participants' attributions from the factors to which they attribute an outcome. For example, studies might define their "attribution measure" as whether a participant claimed ability or effort as the cause (Weiner et al., 1971). In these vignettes, ability is coded by the experimenter as internal and stable, while effort is controllable and unstable. In the former factor, the explicit focus is on locus and stability, but there is a lack of clarity regarding the controllability dimension. This introduces an issue because it's possible that, in a particular study sample, participants have a fixed mindset and believe that they have no control over their inherent ability. As a result, they may exert low effort and quit. However, another sample may have different beliefs about ability, such that they view it as a skill which can be developed; would participants still exhibit low effort and quit? This exemplifies a line of work by Carol Dweck and colleagues, which examines how people's implicit beliefs about ability and intelligence influence their attributions for an outcome (Dweck & Leggett, 1988). Namely, we could have a single cause (i.e., effort) which can result in a multitude of interpretations and outcomes.

This issue underscores our previous point: studies need to include more precise and explicit measures of people's attributions, not just observe the factor to which an individual is attributing. More information or control is required for the way they map the causal dimensions onto that factor. And an even greater issue is that the attributional literature is not always precise in its own interpretations of the three dimensions. For example, the same study by Dweck and Leggett (1988) argues the important connection between beliefs and attributions, highlighting that behavioral output depends on whether participants believe that their intelligence and abilities are *stable or unstable*. However, this does not align with the definition of stability, which tracks whether a factor is transient or constant. A growth mindset is not the belief that our intelligence is transient and fluctuates moment to moment. Rather it is the belief that our intelligence is constant but is amenable to our control through concentrated effort. Such conflations between the stability and controllability dimensions, as well as conflations between locus and controllability dimensions mentioned earlier, lead to a difficulty in interpreting and synthesizing experimental results. Studying attributional dynamics rather than single attributions highlights these nuances in experimental design and interpretation, since a person can attribute an outcome to the same factor multiple times, but the way they conceptualize its causal structure can vary across time (i.e., a lack of motivation can be a transient, controllable state or a stable, uncontrollable clinical condition).

The study of attributional dynamics highlights not only methodological limitations of prior work, but also exciting potential avenues for future research. At first glance the behavioral findings seem obvious: if a person keeps failing at the same task, they will eventually quit. However, we found an interesting visual trend in the data such that participants' rating that Nika would try to change her strategy, skills, or environment peaked in the first (and occasionally second) vignette and decreased steadily over the subsequent vignettes. This suggests that participants initial belief is that Nika will exert increased efforts and act to achieve the goal. It is only following multiple failures and strategy changes that participants believe Nika will quit. There is well-known individual variation in this pattern when it comes to real-world behaviors, for example studies find that low self-efficacy can lead to much earlier quitting, without an initial surge in behavioral adaptation (see Bandura & Cervone, 1983). This finding highlights an impactful area of research which can investigate variation in behavioral adaptation, quitting, and changes in participants' self-related beliefs and motivation throughout trials. Importantly, they could investigate what kinds of attributions underlie these changes, and whether the proposed relationships between attributions and their outcomes hold.

It would be important for future work to compare attributional dynamics when participants make attributions about others' failure versus their own. Particularly, it would be valuable to examine how cultural, individual, and demographic variables influence 1) differences in how participants interpret the attributional structure of a cause across time (e.g., do some cultures construe motivation as uncontrollable versus controllable? How does this influence the way they interpret new information across time?) 2) how likely participants are to make certain attributions in general (ex. how likely participants are to pay attention to causes that are controllable versus uncontrollable, for example one's effort versus one's environment) and how this influences the way they selectively seek out data for causal inference over time. Gaining understanding of these nuances could reveal how people develop beliefs about their affordances in any environment across time and highlights a rich new area for intervention.

The goal of this study was to lay the groundwork in the study of attributional dynamics, by demonstrating that these dynamics do exist, and have the potential for significant impact over and above one-shot attributions. Attributions and outcomes change over time, however the details of those interactions and the relevant factors at play are yet to be explored. We have proposed several directions of research which may prove fruitful; we hope that this proposal leads to renewed interest in the field of attribution theory, whose impact on psychological theory and application has been seminal since its conception, yet whose potential has only begun to be tapped.

References

Abramson, L. Y., Seligman, M. E. P., & Teasdale, J. D. (1978). Learned helplessness in humans: Critique and reformulation. *Journal of Abnormal Psychology*, *87(1)*, 49–74.

Bandura, A., & Cervone, D. (1983). Self-evaluative and self-efficacy mechanisms governing the motivational effects of goal systems. *Journal of personality and social psychology*, *45(5)*, 1017. https://doi.org/10.1037/0022-3514.45.5.1017

Bernstein, W. M., Stephan, W. G., & Davis, M. H. (1979). Explaining attributions for achievement: A path analytic approach. *Journal of Personality and Social Psychology*, *37(10)*, 1810.

Brewin, C. R. (1985). Depression and causal attributions: What is their relation? *Psychological Bulletin*, *98(2)*, 297–309.

Collins, B. E., Martin, J. C., Ashmore, R. D., & Ross, L. (1973). Some dimensions of the internal-external metaphor in theories of personality. *Journal of personality*.

Dweck, C. S., & Leggett, E. L. (1988). A social-cognitive approach to motivation and personality. *Psychological review*, *95(2)*, 256. https://doi.org/10.1037/0033-295X.95.2.256

Frieze, I. H., & Weiner, B. (1971). Cue utilization and attributional judgments for success and failure. *Journal of Personality*, 39(4), 591–606.

Heider, F. (1958). The psychology of interpersonal relations. *John Wiley & Sons*. https://doi.org/10.1037/10628-000

Jones, E. E., & Harris, V. A. (1967). The Attribution of Attitudes. *Journal of Experimental Social Psychology*, 3(1), 1–24. https://doi.org/10.1016/0022-1031(67)90034-0

Kelley, H. H. (1973). The processes of causal attribution. *American Psychologist*, 28(2), 107–128. https://doi.org/10.1037/h0034225

Lau, R. R. (1984). Dynamics of the attribution process. *Journal of Personality and Social Psychology*, 46(5), 1017–1028. https://doi.org/10.1037/0022-3514.46.5.1017

Lewin, K., Dembo, T., Festinger, L., & Sears, P. S. (1944). Level of aspiration. In J. M. Hunt, *Personality and the behavior disorders* (pp. 333–378). Ronald Press.

Martinko, M. J., & Thomson, N. F. (1998). A synthesis and extension of the Weiner and Kelley Attribution models. *Basic and Applied Social Psychology*, 20(4), 271–284. https://doi.org/10.1207/s15324834basp2004_4

Martinko, M. J., Harvey, P., & Dasborough, M. T. (2011). Attribution theory in the organizational sciences: The road traveled and the path ahead. *Academy of Management Perspectives*, 25(2), 81–89.

Moore, B. S., Sherrod, D. R., Liu, T. J., & Underwood, B. (1979). The dispositional shift in attribution over time. *Journal of Experimental Social Psychology*, *15(6)*, 553-569.

Peterson, C., Semmel, A., von Baeyer, C., Abramson, L. Y., Metalsky, G. I., & Seligman, M. E. P. (1982). The attributional style questionnaire. *Cognitive Therapy and Research*, *6(3)*, 287–300.

Seligman, M. E., Abramson, L. Y., Semmel, A., & Von Baeyer, C. (1979). Depressive attributional style. *Journal of abnormal psychology*, *88(3)*, 242.

Weiner, B. (1979). A theory of motivation for some classroom experiences. *Journal of educational psychology*, *71(1)*, 3.

Weiner, B. (1983). Some methodological pitfalls in attributional research. *Journal of educational psychology*, *75(4)*, 530.

Weiner, B. (1985). An attributional theory of achievement motivation and emotion. *Psychological review*, *92(4)*, 548.

Weiner, B., Frieze, I., Kukla, A., Reed, L., Rest, S., & Rosenbaum, R. M. (1987). Perceiving the causes of success and failure. In E. E. Jones, D. E. Kanouse, H. H. Kelley, R. E. Nisbett, S. Valins, & B. Weiner (Eds.), Attribution: Perceiving the causes of behavior. Lawrence Erlbaum Associates, Inc.

Weiner, B., Russell, D., & Lerman, D. (1978). The cognition-emotion process in achievement-related contexts. *Journal of Personality and Social Psychology*, *36(11)*, 1211–1220.

Wells, G. L. (1982). Attribution and reconstructive memory. *Journal of Experimental Social Psychology*, *18(5)*, 447–463. https://doi.org/10.1016/0022-1031(82)90065-8