UC Merced

Title
Toward a normative theory of (self-)management by goal-setting

Permalink
https://escholarship.org/uc/item/5843d1jr

Journal

Authors
Singhi, Nishad
Mohnert, Florian
Prystawski, Ben
et al.

Publication Date
2023

Peer reviewed

Toward a normative theory of (self-)management by goal-setting

Nishad Singhi (nishad.singhi@tuebingen.mpg.de)
Max Planck Institute for Intelligent Systems, Tübingen, 72076, Germany

Florian Mohnert (mail@florian-mohnert.de)

Ben Prystawski (benpry@stanford.edu), Department of Psychology, Stanford University, CA, USA

Falk Lieder (falk.lieder@tuebingen.mpg.de)
Max Planck Institute for Intelligent Systems, Tübingen, 72076, Germany

Abstract

People are often confronted with problems whose complexity exceeds their cognitive capacities. To deal with this complexity, individuals and managers can break complex problems down into a series of subgoals. Which subgoals are most effective depends on people's cognitive constraints and the cognitive mechanisms of goal pursuit. This creates an untapped opportunity to derive practical recommendations for which subgoals managers and individuals should set from cognitive models of bounded rationality. To seize this opportunity, we apply the principle of resource-rationality to formulate a mathematically precise normative theory of (self-)management by goal-setting. We leverage this theory to computationally derive optimal subgoals from a resource-rational model of human goal pursuit. Finally, we show that the resulting subgoals improve the problem-solving performance of bounded agents and human participants. This constitutes a first step towards grounding prescriptive theories of management and practical recommendations for goal-setting in computational models of the relevant psychological processes and cognitive limitations.

Keywords: goal-setting; problem-solving; bounded rationality; computational modeling; management

Introduction

Many complex problems require planning many steps ahead. People often struggle with such problems because their capacity for planning is limited (Prystawski et al., 2022). To overcome this challenge, individuals, managers, and educators often break complex problems down into a series of subgoals (Catrambone, 1998; Drucker, 2012; Simon, 1975).

In principle, it should be possible to use models of human planning (e.g., Callaway et al., 2022), goal-pursuit(e.g., Prystawski et al., 2022), and problem-solving (Anderson, 2013b; Newell & Simon, 1972, e.g.,) to predict which subgoals are most beneficial for people. This creates an untapped opportunity to derive practical recommendations for which subgoals managers and individuals should set from cognitive models of bounded rationality (Gershman et al., 2015; Lewis et al., 2014; Lieder & Griffiths, 2020). Here, we formulate a mathematically precise normative theory of (self-)management by goal-setting that makes it possible to derive optimal subgoals from computational models of bounded rationality (Lieder & Griffiths, 2020). The basic idea is that goal-setting serves to decompose a complex problem into a series of simpler problems that require less planning. To apply this theory, we combined it with a computational model of human goal-pursuit and an optimization algorithm.

This approach allowed us to improve the problem-solving performance of bounded agents and crowd workers in simulations and an online experiment, respectively. These findings suggest that it might be possible to ground prescriptive theories of managerial and personal goal-setting in computational models of bounded rationality.

The outline of this paper is as follows: We first introduce the relevant theoretical background. We then propose our normative theory of goal-setting for (self-)management. The following two sections evaluate our theory in simulations and an online experiment. We close by discussing our results, their implications, limitations, and directions for future work.

Background

Computational models of bounded rationality

Human cognition is constrained by bounded cognitive resources and having to solve complex problems in a limited amount of time (Lieder & Griffiths, 2020). To deal with this, the brain uses heuristic strategies that can perform reasonably well with limited resources. The framework of resource rationality (Lieder & Griffiths, 2020) formalizes this intuition and seeks to understand human behavior as the optimal use of limited cognitive resources and information. Formally, the resource-rational strategy that people should use in an environment E is

$$\psi^\star = \arg\max_{\psi \in \mathcal{S}_B} \mathbb{E}[\mathrm{RR}(\psi, E, B)]$$

where $\mathcal{S}_B$ is the set of all strategies that can be implemented by brain B and the resource rationality RR of a strategy ψ implemented by brain B in an environment E is

$$\mathrm{RR}(\psi, E, B) = \mathbb{E}_{P(\mathrm{result}|s_0, \psi, E, B)}[u(\mathrm{result})]$$
$$- \mathbb{E}[\mathrm{cost}(t_\psi, \rho)|\psi, s_0, B, E]$$

where $u(\mathrm{result})$ is the person's subjective utility of the result obtained by using the strategy ψ in situation s_0 and $\mathrm{cost}(t_\psi, \rho)$ is the total opportunity cost of investing resources ρ used by strategy ψ for time t_ψ.

Resource-rational models take into account cognitive limitations and perform as well as possible under those constraints. This framework has been successfully applied to various cognitive processes such as planning (Callaway et al.,

748

2022), goal pursuit (Prystawski et al., 2022), and decision-making (Bhui et al., 2021).

Simulated Microworlds

Simulated Microworlds (SMWs) are used to study problem-solving in naturalistic, dynamic environments (Funke, 1993). They consist of various variables that interact according to a dynamical system with discrete time steps. The agent can directly manipulate a subset of the variables while only indirectly influencing the values of other variables. For instance, the ultimate aim of the owner of a bakery is to maximize profits. However, she cannot directly control her profit, nor can she influence other factors like demand and sales. Instead, she has to decide how much to spend on advertising, rent, wages, raw materials, etc. in a way that would lead to maximum profit. This makes SMWs a suitable paradigm to study problem-solving in the real world.

In this study, we adopt the SMW introduced by Mohnert et al. (2019), in which participants manage a farm. Their goal is to bring various crops (also referred to as 'states') close to specified target values by adjusting how much fertilizer and pesticides (also referred to as 'actions') are deployed on the fields (Fig. 1). In this SMW, participants have complete knowledge of the relationships between all variables, and suboptimality in their behavior can therefore be attributed to their limited cognitive resources. The dynamics of the system are described by the following equation:

$$\mathbf{s}_{t+1} = \mathbf{f}(\mathbf{s}_t, \mathbf{a}_t) = \mathbf{A}\mathbf{s}_t + \mathbf{B}\mathbf{a}_t$$

where $\mathbf{s}_t \in \mathbb{R}^N$ is the *state* containing values of various crops at time t, $\mathbf{a}_t \in \mathbb{R}^M$ is the *action* consisting of the amounts of various resources used at time t, $\mathbf{A} \in \mathbb{R}^{N \times N}$ determines how various crops affect themselves and each other from time t to $t+1$, and $\mathbf{B} \in \mathbb{R}^{N \times M}$ determines how resources affect crops.

Resource Rational Models of Goal Pursuit

In simulated micro-worlds, participants often struggle to achieve the task's ultimate goal (Funke, 1993; Prystawski et al., 2022) because doing so requires planning many steps ahead and taking various factors into account, which can be cognitively demanding. To effectively model how people pursue goals in SMWs, it is important to consider these cognitive limitations. To this end, Prystawski et al. (2022) developed resource-rational models of goal pursuit, which accommodate limits on people's attention and how many steps they can plan ahead. They found that the model that best explained the problem-solving behavior of the largest proportion of participants in the SMW described above was a hill-climbing model inspired by Newell and Simon's foundational research on human problem-solving (Newell & Simon, 1972).

The Hill-Climbing model This model assumes that the agent's limited computational resources prevent it from planning more than one step into the future. The agent, therefore, chooses the combination of inputs (actions) that maximally reduces the distance to the goal in the very next round, while also minimizing the cost of those inputs. Any input can be described in terms of the direction of the change in the system's state and the distance between the previous state and the next state. Since the aim of the agent is to minimize its distance from the goal, it moves in the direction opposite to the gradient of this distance. Concretely, the action chosen by the agent is:

$$\mathbf{a}_t = -\lambda \cdot \lambda_{opt} \cdot \nabla_{\mathbf{a}} ||\mathbf{f}(\mathbf{s}_t, \mathbf{a}) - \mathbf{g}||_2$$

where λ_{opt} is the optimal step size[1] in the direction of the negative gradient, $\mathbf{g}$ is the goal, $\mathbf{f}(\mathbf{s}_t, \mathbf{a})$ is the state of the agent after taking action $\mathbf{a}$ in the state $\mathbf{s}_t$. The gradient is evaluated at $\mathbf{a} = 0$. λ is a free parameter that captures people's tendency to take steps that are systematically smaller or larger than the optimal step size. Stochasticity in people's actions is captured via noise added to the distance and the direction of the agent's actions (see Prystawski et al. (2022) for details).

The resource-rational perspective on goal-setting

It has been proposed that the function of goals is to reduce the amount of planning that is necessary to reach good decisions (Lieder & Griffiths, 2020). According to a recent refinement of this perspective (Correa et al., 2020; Correa et al., 2022), an optimal sequence of subgoals should minimize the sum of the costs of the actions the person will take and the mental effort they have to invest into planning to select those actions.

Goal-setting in management and self-regulation

A common approach to management is *management by objectives* (Drucker, 2012). In this approach, the manager's first responsibility is to break down a complex problem into a series of subgoals (*objectives*) that their subordinates can achieve efficiently. The manager assigns their subordinates one subgoal at a time. The employees then work towards the assigned subgoal, and once they accomplish it, the manager assigns them the next subgoal. The purpose of this management practice is to enable teams to achieve challenging long-term goals whose achievement requires considerable amounts of planning and problem-solving. Numerous studies have consistently found that organizational productivity benefits from management by objectives in general (Rodgers & Hunter, 1991) and its goal-setting component in particular (e.g., Mento et al., 1987).

Moreover, people can also use goal-setting to manage themselves, improve their own performance, and help themselves achieve their long-term goals (Carver & Scheier, 2001; Latham & Locke, 1991; Zimmerman, 2012).

A normative theory of goal-setting for (self-)management

The finding that people routinely improve the performance of others (or themselves) through goal-setting raises several

---

[1]The optimal step size is the step size that minimizes the distance from the goal immediately after a single step in the optimal direction. It is obtained by setting the derivative of the distance from the goal equal to zero while keeping the direction fixed.

interesting, interrelated questions about which subgoals are most effective, what makes them so effective, and how supervisors managing teams (and individuals managing themselves) can generate them. Here, we approach this question from the perspective of rational analysis (Anderson, 2013a) by formalizing the problem people managing others or themselves solve by setting objectives. For the ease of reading and understanding, we describe the theory for the case of managerial goal-setting. But, in principle, the theory also applies when the person being managed is the manager herself.

Based on the research on goal pursuit summarized in the previous section (Prystawski et al., 2022), a key problem that managers have to solve is that the path to achieving the organization's ultimate goals may be very long and complex relative to individual employees' capacity and/or propensity for planning. From this perspective, managerial goal-setting serves to reduce the amount of planning that is required for effective goal pursuit (cf. Correa et al., 2020; Correa et al., 2022). This, in turn, increases employees' performance on problems that are beyond the cognitive capacity of any single individual. From this perspective, the function of managerial goal-setting is to translate a long-term goal into a sequence of subgoals such that when employees devote their limited cognitive resources to the pursuit of those subgoals they will make more progress towards the long-term goal than they would if they pursued the long-term goal directly.

To formalize this idea, we model employees' decisions using the recently developed model of boundedly rational goal pursuit introduced above (Prystawski et al., 2022). Concretely, we model the employee as an agent α that interacts with an environment E by taking action $\mathbf{a}_t \in \mathbb{R}^M = (a_t^0, a_t^1, \ldots, a_t^M)$ at time t based on the environment's state $\mathbf{s}_t \in \mathbb{R}^N = (s_t^0, s_t^1, \ldots, s_t^N)$. A goal g is characterized by its target values $\mathbf{s}_g \in \mathbb{R}^N$, threshold $\delta_g \in \mathbb{R}$, and scale $\gamma_g \in \mathbb{R}^{N \times N}$. The scale captures the possibility that it is more important to bring some state variables closer to their target values than others, and the threshold specifies how close the agent needs to be to the target values to successfully reach the goal. The agent is considered to have achieved the goal if $\sqrt{(\mathbf{s} - \mathbf{s}_g)^\top \gamma_g (\mathbf{s} - \mathbf{s}_g)} \leq \delta_g$, where γ_g is a diagonal matrix with its ith value being equal to the scale of the ith state variable. Because values of the scale parameter can be hard to interpret for humans, we converted them to tolerance values for human participants in our experiment. The tolerance θ_i for the ith state variable is computed as $\theta_i = \frac{\delta_g}{\sqrt{N \cdot (\gamma_g)_i}} (1)$, where $(\gamma_g)_i$ is the ith diagonal element of γ_g (see Fig. 1).

Subgoals are defined similarly, but for a subset of state variables. Concretely, a subgoal ε can be defined for a set of state variables $D_\varepsilon = (i_1, i_2, \ldots, i_d)$ with target values $\mathbf{s}_\varepsilon$, scale γ_ε, and threshold δ_ε (Fig. 2). The condition for successfully achieving a subgoal is $\sqrt{(\mathbf{s}_r - \mathbf{s}_\varepsilon)^\top \gamma_\varepsilon (\mathbf{s}_r - \mathbf{s}_\varepsilon)} \leq \delta_\varepsilon$, where $\mathbf{s}_r$ is the *reduced state* which is obtained by considering the state variables of $\mathbf{s}$ that are included in D_ε, i.e., $\mathbf{s}_r = (s^i | i \in D_\varepsilon)$. Scales for subgoals can be converted to tolerances as done in Eq. 1 by replacing N with the size of D_ε, δ_g with δ_ε, and γ_g with γ_ε.

The agent starts at $t = 0$ from state $\mathbf{s}_0$ and tries to achieve a series of subgoals $\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_k$ and subsequently the final goal g. Let $G = (\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_k, g)$. The trajectory τ of the agent is $\tau = (\mathbf{s}_0, \mathbf{a}_0, \mathbf{s}_1, \mathbf{a}_1, \ldots, \mathbf{s}_{T-1}, \mathbf{a}_{T-1}, \mathbf{s}_T)$, where T is the duration of the trial. The expected performance of the agent α is the expected value of the quality $\phi(\tau, g, E)$ of potential trajectories τ across all trajectories that might occur, that is

$$\phi_{\mathbf{s}_0}^G(\alpha) = \mathbb{E}_{P(\tau | \mathbf{s}_0, G, \alpha, E)} [\phi(\tau, g, E)]. \quad (2)$$

where $\phi(\tau, g, E)$ measures the performance of the agent during trajectory τ while pursuing goal g in environment E.

In principle, more subgoals would lead to better performance. But, in practice, the desired number of subgoals (k) is limited by the manager's time, the frequency of the manager's communication with the employees, and the employee's need for autonomy (Ryan & Deci, 2006).

With all of these definitions in place, we can now define the optimal solution to the problem of managerial goal-setting as selecting the sequence of subgoals $\varepsilon_1^\star, \varepsilon_2^\star, \ldots, \varepsilon_k^\star$ that maximizes the employee's expected performance given uncertain knowledge about its capacities ($P(\alpha)$) as

$$\varepsilon_1^\star, \varepsilon_2^\star, \ldots, \varepsilon_k^\star = \underset{\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_k}{\arg\max} \mathbb{E}_{P(\alpha)}[\phi_{\mathbf{s}_0}^G(\alpha)]. \quad (3)$$

If the normative theory of managerial goal-setting formalized in Equation 3 is potentially useful, then subgoals derived from this theory should improve worker's performance. In the following two sections, we test this prediction with simulated and real workers, respectively.

Improving the performance of bounded agents

We test our normative theory of managerial goal-setting in the SMW described in the following paragraph. In this section, we apply the theory to compute one optimal subgoal for a resource-rational model of goal-pursuit and check if it improves the model's problem-solving performance.

The management problem: maximizing the productivity of a farm The simulated micro-world used in the present study simulates the problem of managing a farm (see Fig. 1). Starting from the state s_0, the goal of the agents is to bring the values of certain crops (corresponding to state, $\mathbf{s}_t$; shown on the right in Fig. 1) close to the target values specified by the manager. Agents can do so over the course of $T = 20$ steps by using various costly resources (corresponding to action, $\mathbf{a}_t$; shown on the left in Fig. 1). Further, agents have complete information about how crops are influenced by resources and each other, as shown by the weighted edges in Fig. 1. In particular, we studied the setting where agents start from $\mathbf{s}_0 = [80, 20, 90, 10, 70]^\top$ and pursue the goal with $\mathbf{s}_g = [0, 0, 0, 0, 0]^\top, \gamma_g = \mathbf{I}_5$ (i.e., identity matrix of size 5), $\delta_g = 50$, and $\delta_\varepsilon = 1$. For a more concrete description, see "Improving the performance of crowdworkers" below.

The performance (ϕ) of the agent is higher the longer it is within the goal region. Since resources are costly, using

more resources leads to lower performance. We formalize this using the *goal-achievement score* (GAS) to measure the performance ϕ of the agent. If the agent achieves the goal for x out of T time steps, and $y = \sum_t ||\mathbf{a}_t||_1$, then GAS is:

$$\text{GAS}(\tau, g, E) = \max(0, w_1 + w_2 \cdot x - w_3 \cdot y) \qquad (4)$$

where w_1, w_2, and w_3 capture the starting endowment, the reward for achieving the goal, and the cost of resources, respectively. Here, we used $w_1 = 0.2, w_2 = 0.3, w_3 = 0.005$ to capture that achieving the goal is most important. The aim of the agent is to maximize its GAS, while the aim of the manager is to provide a sequence of subgoals to the agent such that pursuing them helps it maximize its GAS.

An important feature of this environment is that the state variable Crowding has an edge with weight = +1.5 starting and ending in it, which creates a positive feedback loop. Without intervention, this feedback loop would cause the value of Crowding to increase exponentially over time. To prevent this, the agent has to bring the value of Crowding close to 0. We therefore predicted that a good subgoal should include Crowding = 0.

Computing an (approximately) optimal subgoal for boundedly rational employees To approximate the optimal goal defined in Equation 3, we approximated the expectation in Eq. 2 by running η noisy simulations of the given agent. We approximated the expectation in Equation 3 by averaging the performances ($\phi_{s_0}^G(\alpha)$ in Eq. 3) over a population (ω) of hill-climbing agents with different values of step size λ that covered the behavior of the largest proportion of participants in Prystawski et al. (2022). To derive ω, we first selected participants from Prystawski et al. (2022) which were best explained using the hill-climbing model, arranged their step sizes (the only free parameter) in ascending order, and selected 30 equally spaced step sizes to cover the entire range of participants. For each agent, we measured the quality of its trajectories by the goal achievement score defined in Eq. 4 (i.e., $\phi(\tau, g, E) = \text{GAS}(\tau, g, E)$).

To further simplify the computational problem, we considered only one 2-dimensional subgoal (i.e., $k = 1$) and used $\eta = 1$. The distance and angular noise in the actions of the hill-climbing agent were drawn from an exponential distribution (with intensity parameter, $\nu = 0.1$) and a von Mises distribution (centered at $0°$ with concentration parameter $\kappa = 40$), respectively.

We then approximated the optimal subgoal defined in Eq. 3, by maximizing the simulated performance using Cross-Entropy (CE) Optimization (De Boer et al., 2005). We ran a separate Cross-Entropy procedure for every possible pair of state variables to optimize for $\mathbf{s}_\varepsilon$ and γ_ε, and then chose the state variables with the highest performance. We ran the CE procedure for 10 iterations with 1000 candidate subgoals in every iteration and selected the top 20% subgoals in every iteration to refine the distribution of potential subgoals.

Results The best subgoal according to our subgoal discovery procedure was Crowding = 0 and SpaceWorms = 4 with the 2×2 diagonal matrix having elements 0.121 and 0.012 respectively as the scale parameter γ_ε. This scale parameter translates to tolerance values of ± 2 and ± 6 for the two subgoal variables, respectively. This is consistent with our prediction that a good subgoal should include Crowding = 0. Simulations with $\eta = 100$ showed that the goal-achievement scores of agents are higher in the presence of subgoals vs. without them according to a Mann-Whitney U-test (0.696 vs. 0.069, $U = 5.9 \times 10^6, p < 0.001$). This highlights the efficacy of the subgoal and our procedure. The standard deviations of the two target values in $\mathbf{s}_\varepsilon$ across 5 runs of the subgoal discovery procedure were 0.71 and 1.41, respectively, which shows that the subgoals discovered by our method are reproducible.

Improving the performance of crowdworkers

We performed a pre-registered experiment to test if the subgoals generated by our method help people in problem-solving. To do so, we tested if providing the subgoal computed in the previous section improves people's performance in the simulated micro-world described above (Figure 1). Participants in the subgoal condition were asked to pursue the subgoal computed by our method before pursuing the final goal (Figure 2), whereas participants in the control condition were directly asked to pursue the final goal (Figure 1). The pre-registration is available at `https://aspredicted.org/5W2_GTV`.

Methods

Participants We recruited 441 crowd workers from the online study platform *Positly*, out of which 234 identified as male, 190 identified as female, and 17 chose not to disclose their gender. The minimum and maximum ages reported were 21 and 76, respectively, with the average age being 40.28 ($SD = 12.05$). Participants spent an average of 39.7 minutes in the experiment. They earned a base pay of \$3 for completing the training. After this, they participated in a practice trial consisting of six rounds ($T = 6$). Participants who achieved the specified goal in this trial were paid \$0.15 and were invited to participate in the main experiment. 302 people participated in the main experiment and received a performance-based bonus, with the average value of the bonus being \$1.28.

Procedure At the beginning of the experiment, participants were shown three instruction videos and were given a chance to participate in three practice trials. Then, they had to take a quiz testing their attentiveness and understanding of SMWs. Following this, they were invited to participate in another practice trial of six rounds ($T = 6$). Participants who achieved the specified goal in this trial were invited to take part in the main experiment. We only analyzed data from the main experiment, which was completed by 302 participants.

We randomly assigned each participant to one of two conditions: the subgoal condition ($n = 150$) and the no subgoal

condition ($n = 152$). Before starting the main experiment, participants in the subgoal condition were informed that they would receive a subgoal that would help them achieve the final goal. Additionally, they were instructed that subgoals can be defined for a subset of crops, and that they should bring the values of these subgoal measures within the specified tolerances of the target values. During the experiment, only the subgoal crops had target values and tolerances next to them (Fig. 2). Participants were given a message upon successfully achieving the subgoal, following which the subgoal disappeared and the final goal was displayed.

In the main experiment, which consisted of one trial of 20 rounds (i.e., $T = 20$), participants earned a bonus payment equal to their goal-achievement score in USD. Participants in both conditions were informed that they would earn a higher bonus by bringing all crops within their target ranges while using as few resources as possible. They were also informed that negative values of resources have the same cost as positive values. The current value of the bonus was displayed on the screen throughout the trial. In addition, the total distance from the subgoal/final goal was also displayed on the screen.

Materials The experiment involved managing a farm on an alien planet (Fig. 1) with the goal of bringing the values of certain crops within the specified tolerances of their target values. The target value, tolerance, and current value of each farming measure were shown alongside it on the screen. Participants could influence the values of crops using various costly resources. They could select the amount (positive or negative) for each resource by either typing in the desired value in the corresponding box or by using the up/down arrow keys. The causal relationships between variables were shown via weighted edges. To reduce cluttering, self-connections were only shown when a variable amplified its value over time (Crowding in Fig. 2).

The starting position and final goal in both conditions were equal to those used previously to compute optimal subgoals. All participants in the subgoal condition were given the subgoal computed by our automatic method: Crowding = 0±2 and SpaceWorms = 4±6. For the practice trials, we employed an SMW which was different from the one used in the main experiment but followed the same rules.

Results

The average goal-achievement scores were 1.33 and 1.23[2] in the subgoal and no subgoal conditions, respectively. This difference was not statistically significant according to a Mann-Whitney U-test ($U = 10595$, $p = 0.11$). However, the proportion of participants with a positive score was significantly higher in the subgoal condition than in the control condition according to a two-proportions z-test (43.33% vs. 31.6%, $z = -2.11$, $p = 0.035$).

---

[2]The improvement in score was robust to changes in the parameters used to compute GAS. Scores were higher in the experimental condition whenever the reward for achieving the goal (w_2) was greater than 0.1, and the cost of resources (w_3) was less than 0.13.
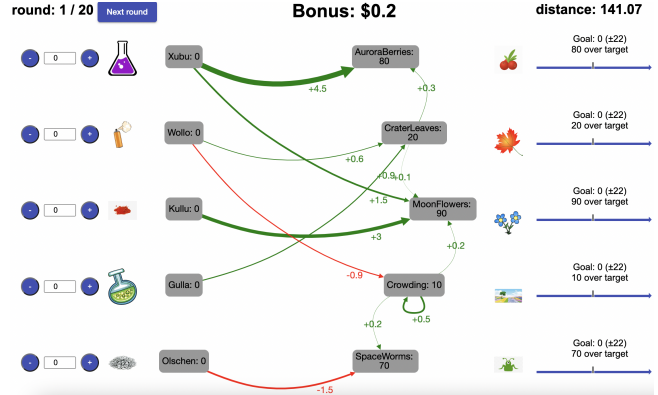

Figure 1: Screenshot of the Simulated Microworld shown to participants in the experiment. Participants could enter the desired values of the resources in the boxes on the left. The values of the crops and their goal values are displayed to the right. The edges represent causal relations between variables.
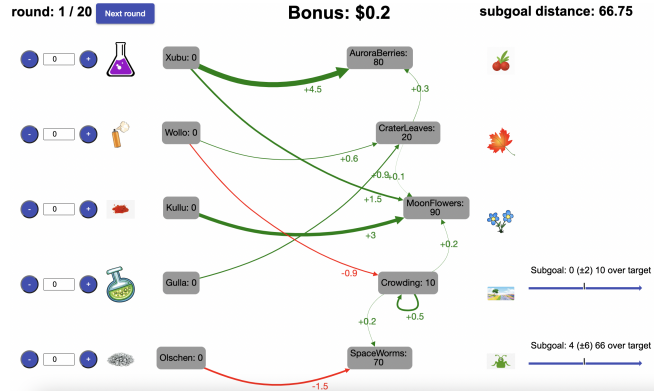

Figure 2: Screenshot of the Simulated Microworld showing the subgoal discovered by our method. Note that crops not in the subgoal do not have corresponding target and tolerance values because the subgoal has not been achieved yet.

The amount of resources used by a participant was computed as $\sum_t ||\mathbf{a}_t||_1$. We found that participants in the subgoal condition used substantially fewer resources than the control condition (494.45 vs. 859.5, $U = 13092$, $p = 0.012$).

Additionally, we also computed participants' distance score as $DS = \sqrt{||\mathbf{s}_T - \mathbf{s}_g||_2^2 + c \cdot \sum_{t=0}^{T-1} ||\mathbf{a}_t||_2^2}$. This score captures how close an agent gets to the final goal at the end of the trial while penalizing it for using more resources. Smaller values of DS are better. Following Prystawski et al. (2022), we used $c = 0.01$. In the subgoal condition, the median value of the distance score was numerically lower than in the control condition (492 vs. 988). However, because of the high variance, this difference was not statistically significant according to a one-sided Mann-Whitney U-test ($U = 12466$, $p = 0.08$). Taken together with the significant reduction in the amount of resources used, this result indicates that sub-

goals can help people achieve their goals more efficiently.

In summary, the subgoal condition performed numerically better than the control condition on all outcome measures, was significantly more likely to achieve a positive goal-achievement score, and used the farm's resources significantly more efficiently. These findings are inconclusive, but broadly consistent with the interpretation that the automatically derived subgoal had a small positive effect on people's performance in the problem-solving task.

Discussion

Goal-setting is commonly used to improve people's ability to solve complex problems (Catrambone, 1998; Drucker, 2012; Locke & Latham, 2002). One of the reasons why goal-setting is effective is that it reduces the amount of planning that is necessary for goal achievement. It should therefore be possible to leverage resource-rational models of planning and goal-pursuit (e.g., Callaway et al., 2022; Prystawski et al., 2022) to improve the theory and practice of improving performance through goal-setting (Locke & Latham, 2002).

To explore this approach, we have proposed a normative theory for (self-)management by goal-setting and applied it to computationally derive subgoals from a resource-rational model of goal pursuit. Our proof-of-concept simulations and experiment suggest that it might be possible to derive helpful goal suggestions from resource-rational models of goal pursuit. This illustrates that it is, at least in principle, possible to ground recommendations for goal-setting in the theory of resource-rationality. Our work could therefore be considered a first step towards establishing empirically supported computational models of bounded rationality as a micro-foundation for prescriptive theories of (self-)management.

The main limitation of the present work is that empirical evidence for our method's ability to improve human problem-solving was mixed. To explain the mixed results, it is worth noting that the self-amplifying dynamics of the simulated micro-world we used in this experiment made participants' scores extremely variable. This variability, in turn, reduced the power of our statistical tests. As a consequence, even large numerical differences were not always statistically significant. Concretely, the high variability of the scores resulted from the presence of a positive feedback loop that caused Crowding to increase exponentially over time. Once Crowding exceeded a certain value, the exponential growth became unstoppable and participants could no longer control the system. This made the task very challenging for participants, even in the presence of the optimal subgoal.

Future work should investigate why our participants benefited less from the provided subgoal than our simulations had predicted, and improve the model (and the resulting subgoal) accordingly. Follow-up experiments should also investigate the moderating role of individual differences in cognitive ability and motivation. Moreover, our assessment of the method was limited to a single problem in just one simulated micro-world. Future experiments should assess the generalizability of our findings to other problems in other environments. Such studies could jointly identify under which conditions the subgoals recommended by our method are most beneficial and who benefits the most.

While previous work explored resource-rationality in planning tasks with discrete states (Correa et al., 2020; Correa et al., 2022), we applied it to goal-setting for complex problem-solving in dynamic environments with continuous states and inputs. Moreover, while previous work made the unrealistic assumption that the goal is always achieved, our normative theory of goal-setting takes into account that goal-achievement is the exception rather than the norm and that maintaining the desired state is also an important part of the problem. Also, our work draws on an evidence-based process model of human goal-pursuit, instead of assuming that people use search algorithms developed for computers.

Moreover, while Correa et al. (2020) and Correa et al. (2022) studied how people do and machines should decompose tasks into subtasks, our goal was to formulate a normative theory of (self-)management. As far as we know, previous research on improving managerial goal-setting did not explicitly engage with mechanistic models of goal pursuit and bounded rationality and did not develop computational methods for computing optimal subgoals.

The work we have presented in this article can be extended in several directions. One direction is to improve the current method for computing optimal subgoals. Possible improvements include generating a series of multiple subgoals, improving the accuracy and/or speed of the optimization algorithm, considering higher-dimensional subgoals, incorporating the mental effort of planning and goal-pursuit into the objective function (cf. Correa et al., 2020; Correa et al., 2022). Future work can also use our normative theory of goal-setting as the starting point for resource-rational analyses of how people set goals for others and themselves, respectively.

Finally, we hope that in the long run, the research begun in this project will improve the theory and practice of individual and organizational goal-setting. As a step in this direction, future experiments should compare the effectiveness of the subgoals generated by our method against the effectiveness of subgoals derived from previously proposed heuristics and subgoals chosen by participants. Another step in this direction could be to use an improved version of our theory to generate optimal subgoals for a wide range of problems, and then characterize what features useful subgoals have in common.

Although research on grounding prescriptive theories in computational models of bounded rationality is still in its infancy, it is at least beginning to suggest that this is possible to leverage the rigorous methods of computational cognitive science to generate practical knowledge and useful technologies for helping people and organizations become more effective (Lieder & Prentice, 2022; Lieder et al., 2022). We hope that future work in this direction will establish a solid cognitive science foundation for helping people, teams, and organizations set better goals.

Acknowledgements

References

Anderson, J. R. (2013a). *The adaptive character of thought*. Psychology Press.

Anderson, J. R. (2013b). *The architecture of cognition*. Psychology Press.

Bhui, R., Lai, L., & Gershman, S. J. (2021). Resource-rational decision making. *Current Opinion in Behavioral Sciences*, *41*, 15–21.

Callaway, F., van Opheusden, B., Gul, S., Das, P., Krueger, P. M., Griffiths, T. L., & Lieder, F. (2022). Rational use of cognitive resources in human planning. *Nature Human Behaviour*, *6*(8), 1112–1125.

Carver, C. S., & Scheier, M. F. (2001). *On the self-regulation of behavior*. Cambridge University Press.

Catrambone, R. (1998). The subgoal learning model: Creating better examples so that students can solve novel problems. *Journal of experimental psychology: General*, *127*(4), 355.

Correa, C. G., Ho, M. K., Callaway, F., & Griffiths, T. L. (2020). Resource-rational task decomposition to minimize planning costs. *arXiv preprint arXiv:2007.13862*.

Correa, C. G., Ho, M. K., Callaway, F., Daw, N. D., & Griffiths, T. L. (2022). Humans decompose tasks by trading off utility and computational cost. *arXiv preprint arXiv:2211.03890*.

De Boer, P.-T., Kroese, D. P., Mannor, S., & Rubinstein, R. Y. (2005). A tutorial on the cross-entropy method. *Annals of operations research*, *134*, 19–67.

Drucker, P. (2012). *The practice of management*. Routledge.

Funke, J. (1993). Microworlds based on linear equation systems: A new approach to complex problem solving and experimental results. In *Advances in psychology* (pp. 313–330). Elsevier.

Gershman, S. J., Horvitz, E. J., & Tenenbaum, J. B. (2015). Computational rationality: A converging paradigm for intelligence in brains, minds, and machines. *Science*, *349*(6245), 273–278.

Latham, G. P., & Locke, E. A. (1991). Self-regulation through goal setting. *Organizational behavior and human decision processes*, *50*(2), 212–247.

Lewis, R. L., Howes, A., & Singh, S. (2014). Computational rationality: Linking mechanism and behavior through bounded utility maximization. *Topics in cognitive science*, *6*(2), 279–311.

Lieder, F., & Prentice, M. (2022). Life improvement science. In F. Maggino (Ed.), *Encyclopedia of quality of life and well-being research*. Springer.

Lieder, F., & Griffiths, T. L. (2020). Resource-rational analysis: Understanding human cognition as the optimal use of limited computational resources. *Behavioral and brain sciences*, *43*.

Lieder, F., Prentice, M., & Corwin-Renner, E. (2022). An interdisciplinary synthesis of research on understanding and promoting well-doing. *Social and Personality Psychology Compass*, e12704.

Locke, E. A., & Latham, G. P. (2002). Building a practically useful theory of goal setting and task motivation: A 35-year odyssey. *American psychologist*, *57*(9), 705.

Mento, A. J., Steel, R. P., & Karren, R. J. (1987). A meta-analytic study of the effects of goal setting on task performance: 1966–1984. *Organizational behavior and human decision processes*, *39*(1), 52–83.

Mohnert, F., Tošic, M., & Lieder, F. (2019). Testing computational models of goal pursuit. *Conference on Cognitive Computational Neuroscience (CCN 2019)*, 1095–1098.

Newell, A., & Simon, H. A. (1972). *Human problem solving* (Vol. 104). Prentice-hall Englewood Cliffs, NJ.

Prystawski, B., Mohnert, F., Tošić, M., & Lieder, F. (2022). Resource-rational models of human goal pursuit. *Topics in Cognitive Science*, *14*(3), 528–549.

Rodgers, R., & Hunter, J. E. (1991). Impact of management by objectives on organizational productivity. *Journal of Applied Psychology*, *76*(2), 322.

Ryan, R. M., & Deci, E. L. (2006). Self-regulation and the problem of human autonomy: Does psychology need choice, self-determination, and will? *Journal of personality*, *74*(6), 1557–1586.

Simon, H. A. (1975). The functional equivalence of problem solving skills. *Cognitive psychology*, *7*(2), 268–288.

Zimmerman, B. J. (2012). Goal setting: A key proactive source of academic self-regulation. In *Motivation and self-regulated learning* (pp. 267–295). Routledge.