Deep Q-Life Sobriety:

Applying RL Habit Loops to Alcohol-Free Living

Dr. Yves J. Hilpisch[1]

January 17, 2026 (preliminary draft)

Abstract

We extend the Deep Q-Life framework into a field guide for quitting alcohol and sustaining sobriety. Readers map cravings as Markov decision processes, craft replacement policies through a Friday-evening case study, and run experience replay with clinicians and accountability pods. Self-coaching scripts, purpose-aligned activity portfolios, and safety protocols feed a phased roadmap from detox to long-term expansion. The conclusion offers a motivating outlook for applying reinforcement learning discipline to lifelong recovery.

[1] Get in touch: `https://linktr.ee/dyjh`. Web page `https://hilpisch.com`. Research, structuring, drafting, and visualizations were assisted by GPT 5.1 as a co-writing tool under human direction.

Contents

1 Why Sobriety, Why Now

This applied companion extends *Deep Q-Life: Markov, Bellman, and Habits* [1] with a sobriety-focused playbook. We assume the reader is familiar with the original framework and now wants to translate those principles into a zero-alcohol lifestyle.

Sobriety decisions often surface after years of living the Bellman paradox: the drink feels rewarding in the short run, yet compounds into escalating negative returns on health, relationships, finances, and identity. Deep Q-Life invites us to treat the pattern explicitly as a Markov decision process (MDP). The state is not "alcohol yes/no", but the nuanced trigger—the stress-laden commute home, the celebratory dinner, the lonely Sunday afternoon. Once that state is named, we can compare actions by total value: does the next hour feel better and does it set up better options tomorrow?

This guide serves three audiences. First, high-functioning professionals who cycle between control and binge but want a hard reset. Second, people who have tried to quit, relapsed, and need a structured way to study the data from each attempt. Third, support partners who coach or mentor someone choosing sobriety. Each reader gets step-by-step prompts grounded in the same reinforcement learning (RL) loops as the main text, but rephrased for triggers that involve alcohol, social pressure, and emotional volatility.

Medical Partnership

Quitting alcohol cold turkey can be medically dangerous. Any reader with heavy or long-term consumption must coordinate with a physician or addiction specialist before changing consumption patterns. Detox protocols, medications, and emergency plans belong in the professional domain. Use this document as a complement to medical care, therapy, and mutual-aid groups, never as a substitute.

Aligned with that partnership is the third pillar: community scaffolding. We pair the RL tooling with accountability groups, coaches, and supportive friends so that policy updates are rehearsed in safe spaces and reinforced by peers. The remainder of the paper builds a playbook for day-zero planning, weekly reviews, and relapse mitigation using the state-action-reward language introduced in the Deep Q-Life framework.

2 Deep Q-Life Theory Refresher

Deep Q-Life compresses the essential building blocks of reinforcement learning into a language that decision-makers can deploy without code. We restate the pillars here with sobriety in mind; readers can revisit the full derivations in the primary paper [1].

State, action, reward. Every craving moment can be modelled as a Markov decision process (MDP) with state s capturing cues (location, people, emotions), action a representing the chosen response, and reward r reflecting the blend of immediate feeling and delayed payoffs. The sobriety challenge is to widen the action set beyond "drink or white-knuckle", adding pre-committed routines, social scripts, and values-aligned activities.

Bellman optimality. The Bellman equations formalize why short-term relief is rarely optimal. We maximize the action-value function $Q^*(s,a) = R(s,a)+\gamma \sum_{s'} P(s' \mid s,a) \max_{a'} Q^*(s',a')$ by choosing actions that pay now and propagate better states forward [2, 3]. In sobriety terms, we seek routines that neutralize cravings while increasing the probability of waking up empowered tomorrow.

Deep Q-learning loop. Sobriety practice becomes an experience replay engine. Each day produces transitions (s, a, r, s') stored in a journal. Weekly reviews approximate the temporal-difference target $y = r + \gamma \max_{a'} Q_{\text{target}}(s', a')$ and update a simplified policy—the collection of scripts, commitments, and environment tweaks that define the next week's choices [4]. This is the human analog of online and target networks.

Identity and environment. Behavioural science reminds us that the winning policy is supported by identity statements ("I am a non-drinker") and environmental design (remove cues, add supports) [5, 6]. The RL model ensures these interventions are evaluated systematically rather than as vague resolutions. Each subsequent section links these pillars to concrete tools: trigger maps, policy experiments, and review rituals.

Soft Success Factors in Deep Q-Life Sobriety

The mathematical backbone of Deep Q-Life assumes consistent observation, stable preferences, and steady training. Human reality is noisier. Certain "soft" capabilities act as multipliers on the technical framework by improving the effective state representation, the reward signal, and the discount factor.

State-awareness and mindfulness. Accurate decisions require accurate sensing. The skill of pausing and naming one's internal state (tired, lonely, restless, triggered) sharpens the observation space that feeds the MDP. Mindfulness techniques—brief breath checks, body scans, and emotion labeling—turn blurred, reactive states into clearer inputs for the policy. A useful image is a headlight in fog: in drinking mode, the beam is narrowed to a thin cone pointed at a single option, "drink now". State-awareness widens the beam to take in the whole intersection of life—moving away, texting a friend, going for a walk, doing twenty-five push-ups, or starting dinner early. Figure 1 sketches this widening field of view: from a narrow tunnel locked on a bottle of beer to a broad scan of all other *feasible actions* that also notices the past (recent streaks), the environment (who else is present), and the people affected (partner, children, colleagues). Figure 7 provides a compact summary as a "360-degree scan + 90-degree foresight" checklist for real-time decisions.

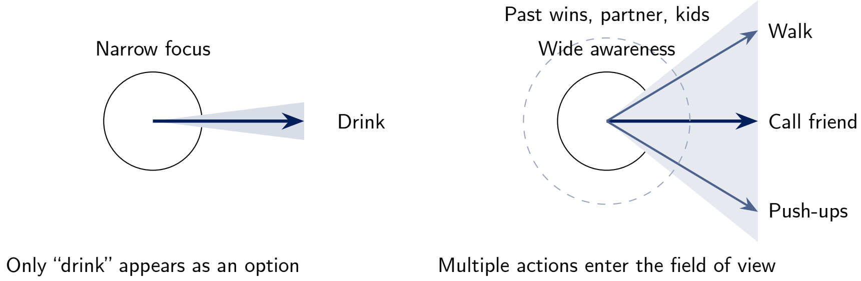

Figure 1: Awareness as widening the field of view. On the left, attention narrows to a thin tunnel where "drink now" is the only visible action. On the right, state-awareness widens the cone so that alternative moves—movement, connection, self-care—and the broader context (past streaks, loved ones) are all taken into account before choosing.

Distress tolerance and urge surfing. In formal terms, "do nothing and observe" is a valid action. In practice, it is only available if a person can tolerate the discomfort of craving without immediately acting. Training urge surfing makes this action feasible: instead of collapsing into drink or white-knuckling, the policy can choose "stay with the wave and ride it out," which often yields higher long-term value.

Future-self continuity. Deep Q-Life uses the discount factor γ to weigh future rewards. Psychologically, this maps to how real and important the future self feels. Exercises that strengthen future-self continuity—imagining tomorrow morning, next month, or a one-year-sober version of oneself—effectively increase the subjective discount factor, raising the value of health, relationships, and purpose. Visually, imagine several gaze angles all starting from the same vantage point. In relapse mode your attention is aimed at a steep angle downward, fixating on the next five minutes. As the angle becomes flatter, the same person can see along the time axis toward tomorrow, and ultimately toward a one-year-sober self on the horizon. Figure 2 illustrates these different viewing angles from a single starting position.

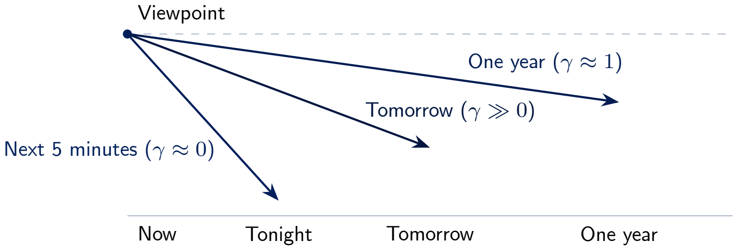
Viewpoint
One year ($\gamma \approx 1$)
Tomorrow ($\gamma \gg 0$)
Next 5 minutes ($\gamma \approx 0$)
Now
Tonight
Tomorrow
One year

Figure 2: Future-self as adjusting the angle of view from a single vantage point. A steep gaze from the horizon cuts down quickly to the next five minutes, while progressively flatter angles extend further along the time axis toward tomorrow and out to a one-year-sober self, increasing the weight of longer-term rewards in day-to-day decisions.

Self-compassion after slips. Learning systems do not throw away the model after a bad sample. Similarly, self-compassion keeps the episode going after a slip. Instead of "I failed, so the policy is useless," the stance becomes "I collected a high-cost data point; what state, action, and support variables need updating?" This prevents catastrophic abandonment of the policy and preserves the long-term gradient.

Values clarity and identity shift. A clear answer to "What kind of person am I becoming?" stabilises the reward function. When sobriety is framed as service to core values—being a present parent, a reliable colleague, a healthy creator—the mapping from outcomes to reward becomes less volatile. Identity statements ("I protect my brain for deep work" or "I am a non-drinker who honours future me") constrain the policy search space in a helpful way.

Figure 3 sketches the feedback loop that anchors the remainder of the guide.

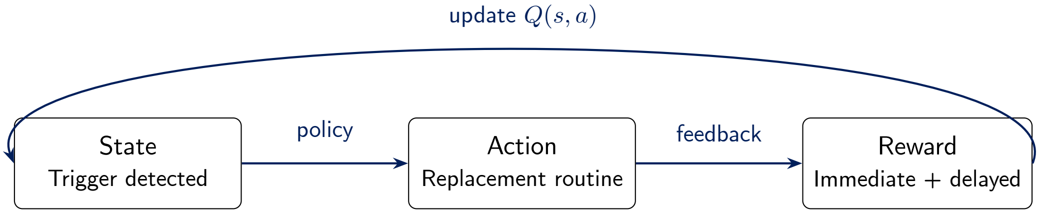

Figure 3: Sobriety-focused reinforcement loop showing how trigger states, chosen actions, and reward evaluations feed the Deep Q-Life update.

3 State, Action, Reward for Alcohol Triggers

Mapping cravings into Markov decision process (MDP) components works best when the process feels tangible. Think of each day as riding a metro line: every station is a trigger state, every train you board is an action, and the eventual neighborhood you end up in is the reward. We want to know which stations make relapse most likely, what alternate trains are available, and how pleasant each destination becomes over time.

Three-Breath State Check

Before choosing an action, pause for three slow breaths. On the first breath, notice and silently name your body state ("tense shoulders, racing heart, low energy"). On the second breath, name the emotion ("anxious, lonely, frustrated, excited"). On the third breath, list at least two non-drinking options you could take next. This turns a blurry craving moment into a clearer state in your MDP and widens the action set beyond "drink vs. resist".

States: naming the stations. Start with a week-long observation sprint. For every urge to drink, log time, location, company, body sensations, and thoughts. Cluster these notes into canonical states such as `commute_fatigue_alone`, `celebration_restaurant_team`, or `sunday_boredom_home`. Describing triggers at this granularity prevents the blanket label "evening" from hiding different dynamics. Figure 4 illustrates how state-specific craving intensity shifts as policies improve.

Actions: swapping trains. In each state, list three categories of responses:

1. **Default drift:** what historically happens (stop at the bar, pour wine, justify "just one").
2. **Micro-barriers:** friction moves that slow the decision (call a sponsor, delay ten minutes, change location).
3. **Value-aligned routines:** deliberate replacements such as heading to a boxing class, prepping a recovery meal, or joining an online support meeting.

Imagine a decision board like a subway map: the default train is a straight line to relapse station, while the replacement lines branch toward identity-aligned destinations. Colour-code the options to visualize which lines you want to board when cravings spike.

Rewards: rating the destination. Assign immediate and delayed scores on a 1–10 scale. Immediate reward captures the momentary shift in stress or pleasure; delayed reward captures sleep quality, relationships, self-respect, and financial impact the next day. For example, ordering a mocktail may score a 4 in immediate relief but rises to an 8 when you factor in waking without shame. Supplement numeric scores with narrative notes ("Woke up clear, went for run"). Figure 8 visualizes the core tradeoff between immediate fun, next-day costs, and higher-value alternatives that compound into peak performance. These values populate the future Q-value estimates that drive policy updates.

Transition probabilities: how likely is the next stop? Just as trains run at scheduled intervals, certain triggers almost guarantee the next state. If `commute_fatigue_alone` leads to `bar_counter` 70% of the time, we need a stronger intervention than if the transition happens only occasionally. Use simple frequency counts from your journals to estimate these probabilities; they feed the Bellman calculations later.

To tie the components together, build a trigger ledger: a spreadsheet or markdown table listing states, likely next states, available actions, and reward estimates. Review it weekly, updating the entries with new evidence. Over time, the ledger becomes the dataset that powers both coaching conversations and self-reflection.

To support these diagnostics we plot craving intensity and policy experiments, as previewed in Figure 4.

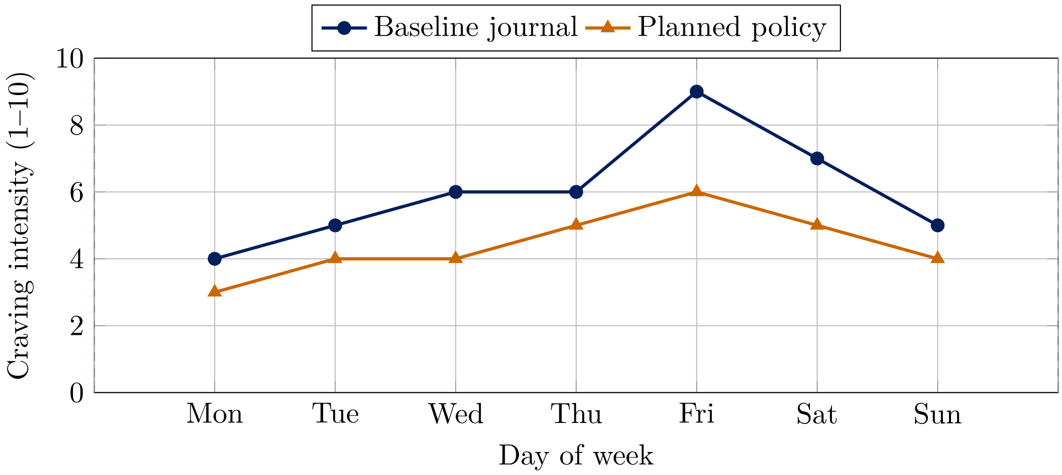

Figure 4: Weekly craving intensity curve comparing baseline self-reports with projected improvements after policy updates. Final data will be sourced from sobriety journals and review sessions.

4 Case Study: Rewriting the Friday Evening Policy

To illustrate the framework, we follow Alex, a product manager who ends every workweek with colleagues at a wine bar. The state we logged in the trigger ledger is `friday_wrapup_team_bar`. Immediate reward scores high because the drink signals closure and bonding, yet delayed reward plummets: Saturday mornings begin with headaches, Sunday brings guilt, and the following week restarts with low energy. Transition analysis shows that leaving the office after 6 p.m. with teammates leads to the bar 80% of the time, making it a prime candidate for a policy rewrite.

Baseline policy. Under the default plan, Alex sticks with the team until the second drink, then catches a late train. Immediate reward is a 7, delayed reward a 2. The action set is narrow, essentially a single edge in the MDP. Journals reveal that cravings spike around 5:15 p.m. while anticipating weekend tasks.

Designing an alternative. We prototype a replacement routine dubbed `friday_transition_ritual`. The new policy inserts a pre-committed call with a mentor at 5 p.m., offers to lead the team debrief during the first round (ordering sparkling water), and then leaves for a scheduled boxing class at 6 p.m. This action combination keeps social connection while redirecting the evening. Immediate reward estimates a 5 (less buzz), but delayed reward climbs to 8 thanks to early sleep, Saturday workouts, and improved relationships at home.

Measuring progress. We deploy a simple log for twelve consecutive Fridays. Each entry records state, action, immediate/delayed reward, and whether the boxing class or bar was chosen. The Q-value estimates in Figure 5 are updated with the actual rewards: the orange curve should decay as the old routine loses appeal, while the navy curve should rise with repeated wins. Weekly reviews also document qualitative notes such as “arrived home energized” or “teammates respected boundary,” strengthening the value signal.

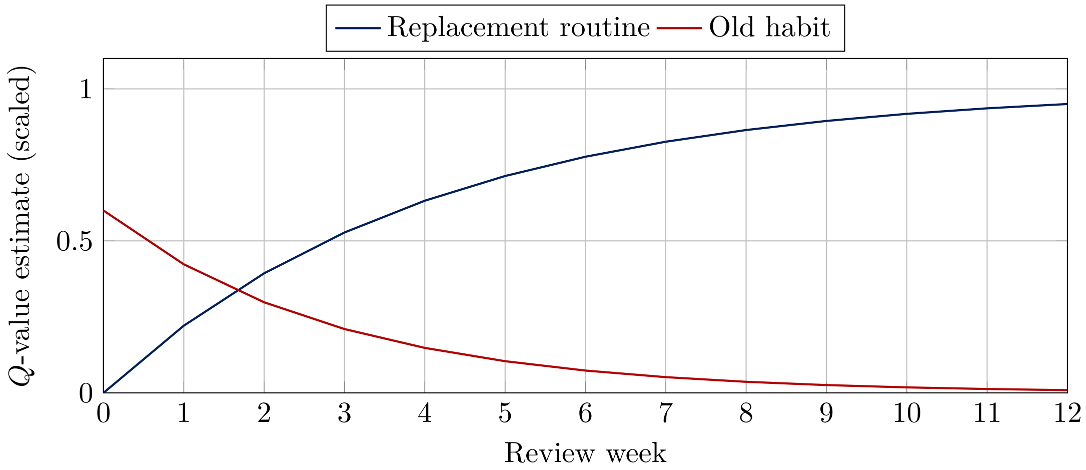

Figure 5: Evolution of estimated Q-values for the competing Friday routines across 12 review weeks. In practice these curves are computed from logged rewards and relapse incidents.

Handling slips. In week four Alex misses the class because of a last-minute release issue and follows the old habit. Instead of framing it as failure, the replay meeting tags the incident as a data point about transition pressure. The policy now includes a contingency: if work extends past 6 p.m., Alex triggers a video check-in with a recovery buddy and reschedules the workout for Saturday morning.

This case demonstrates how the abstract Bellman equation becomes a human story: we map states, diversify actions, quantify rewards, and iteratively adjust the policy. The same structure can model other high-risk windows such as Sunday brunch or business travel.

5 Experience Replay and Support Systems

Deep Q-Life sobriety thrives on deliberate practice loops, not willpower. Think of your week as a flight data recorder: every trigger, action, and consequence gets captured so you can replay the footage with your ground crew and tune the autopilot.

Mindful Replay Minute

Before a weekly review, sit for one minute with feet on the floor and hands resting on your legs. Breathe in for four counts, out for six, and notice any urge to judge the week as “good” or “bad”. Let the judgment pass and instead ask, “What did this week teach me about my policy?” Enter that sentence as the first line of your review. This gentle reset improves the signal-to-noise ratio in your replay buffer.

Daily logging: the replay buffer. Build a lightweight template (journaling app, spreadsheet, or printed card) with the columns `time`, `state code`, `action chosen`, `immediate reward`, `delayed reward`, `support used`. Keep entries concise—two sentences or a few numbers. If you already track sleep, heart rate, or mood, append those so that physiological signals become part of the learning dataset.

Weekly review ritual: target network update. Set a fixed slot (e.g., Sunday evening) to read the week’s entries. Use highlighters or tags to flag successful substitutions, partial wins, and slips. Calculate simple aggregates: average craving intensity, count of meetings attended, hours of restorative sleep. Summarize the top three lessons into an “action memo” that becomes the policy for the upcoming week. This mirrors the target network update in DQL—slow, steady adjustments rather than impulsive pivots.

Accountability pods: distributed replay. Share the memo with at least one support partner. Options include:

- **Therapy or coaching sessions:** structured space to interpret data, surface blind spots, and adjust values scales.
- **Mutual-aid meetings (AA, SMART Recovery):** opportunities to test scripts aloud, collect peer strategies, and celebrate streaks.[2]
- **Digital cohorts:** text groups or forums where members post daily check-ins; treat these as external buffers to prevent silent escalation.

Each partner sees a curated summary, not raw diaries, respecting privacy while reinforcing alignment.

Callout dashboards. Use callout boxes inside the manuscript to host ready-to-print trackers: a seven-day urge log, a "support contact tree," or a relapse response flowchart. During layout, integrate QR codes that link to editable copies if desired.

Escalation protocol. Define thresholds that trigger immediate outreach (e.g., more than two high-intensity cravings in a day, skipped medication, or intrusive thoughts). This keeps the process from drifting into mere journaling; it becomes an operational system with alarms.

By treating experience replay as a shared, data-rich practice, sobriety stops being a lonely fight and becomes a collaborative learning project. The metrics you collect here feed the value curves in Figure 5 and the dashboards introduced later in Figure 6, ensuring that every tweak is evidence-based.

6 Self-Coaching Scripts and Cognitive Reframes

Self-coaching is the cockpit voice that guides you when no sponsor or therapist is on the line. We cultivate it the way pilots rehearse checklists before takeoff: practice the lines in calm air so they surface automatically during turbulence.

Urge Surfing Script

When a craving hits, imagine it as a wave moving through your body. For 90–180 seconds, track where you feel it most (throat, chest, stomach) and silently repeat, "This is a wave; waves rise and fall; I do not have to act." Notice how the intensity changes without taking a drink. In state–action terms, you are practising the "observe and ride it out" action so that it becomes available when the stakes are higher.

From craving chatter to mission control. Begin by transcribing the exact phrases that arise before a drink—"I deserve a break," "One glass won't matter," "They'll think I'm rude if I pass." Next to each line, craft an intentional counter-script rooted in your values and data. For example, "I deserve a break" becomes "My future self deserves a clear Saturday, and walking out now secures it." Rehearse these scripts aloud while visualising the trigger, the same way athletes run mental plays before a match.

[2] AA stands for Alcoholics Anonymous, a peer-support fellowship built around the Twelve Steps; SMART Recovery is Self-Management and Recovery Training, a science-based mutual-aid community emphasising cognitive-behavioural tools and secular meetings. Both offer structured accountability that complements clinical care.

Belief audits. Once a week, conduct a "belief audit": list the top three thoughts that fueled cravings, challenge their accuracy, and replace them with statements that align with your non-drinker identity. Think of it as refactoring legacy code—we identify buggy logic and rewrite the function so it no longer crashes under load. Capture the new beliefs on cards or in a notes app for quick review.

Scenario simulations. Create mini-scripts for high-risk scenarios (office party, family conflict, celebration). Each script includes: trigger description, grounding action (breath, posture shift), primary line, backup line, exit plan. Role-play these with a friend or record yourself delivering them; the goal is to make the language familiar and the body memory strong.

Compassionate debriefs. After any slip or surge in craving, run a compassionate debrief. Instead of "I failed," frame it as "Mission log: system overload at 20:30, cause identified as skipped meal, new protocol—eat before commute." This keeps the tone scientific and self-supporting, avoiding shame spirals that can trigger abandonment of the policy.

7 Replacement Activities and Purpose Alignment

Sobriety is not just about avoiding one behaviour; it is about filling the reclaimed time with choices that reinforce who you are becoming. Picture your week as a wardrobe. Removing alcohol is equivalent to tossing a frayed jacket—the closet now needs tailored garments that fit the life you want.

Future-Self Check-in

Before a high-risk evening, close your eyes for two minutes and picture three versions of yourself: tomorrow morning, three months from now, and one year sober. Ask each, "What action tonight would you most thank me for?" Capture the answers in a sentence or two, then choose the replacement activity that best serves those future selves. This simple visualization nudges your internal discount factor γ upward by making long-term rewards feel more immediate.

Build a replacement portfolio. Assign each day a mix of four activity classes:

Physical reset High-intensity exercise, yoga, cold plunges, or long walks. These discharge accumulated stress hormones and prove to your nervous system that the body can generate its own dopamine.

Creative expression Writing prompts, music practice, sketching, or cooking experiments. These reintroduce flow states that mirror the novelty once sought in drinks.

Service and connection Volunteering, mentoring, or family rituals. Serving others creates rewards beyond self-gratification and strengthens community ties.

Restorative stillness Guided breathing, mindfulness, or intentional rest. This class ensures recovery periods so you do not sprint into burnout and trigger cravings.

Treat the portfolio like an investment mix—adjust the weights based on mood data, energy levels, and the upcoming demands highlighted in your replay logs.

Purpose mapping. Translate each activity into a value statement. For instance, "Boxing class" supports the value "strength and discipline," while "Sunday breakfast with partner" supports "intimacy." Capture these mappings in a simple table so that every scheduled block carries meaning, not just distraction. When cravings arise, review the table and choose the action that best aligns with the value you want to reinforce that moment.

Micro-rewards. Pair activities with immediate reinforcers to compete with alcohol's quick payoff: savoury post-workout smoothies, stickers on a progress board, or a message to your accountability pod celebrating the choice. Feeding these signals back into your reward ledger keeps the Q-values honest.

Tracking fit. At the end of each week, rate the portfolio on mood, energy, social connection, and progress toward personal goals. These scores feed into the roadmap metrics later visualised in Figure 6. Activities that consistently rate low either need refinement or retirement—just as you would rebalance an investment fund that underperforms.

By curating replacement activities around purpose rather than mere distraction, the sober lifestyle becomes self-reinforcing: every choice deposits value into the identity you are constructing.

8 Safety, Relapse Plans, and Professional Guidance

Sobriety work is inseparable from safety planning. Alcohol withdrawal can trigger seizures, delirium tremens, or other medical emergencies; any reader with moderate to severe use must consult a physician before altering dosage. Capture medical instructions, medication schedules, and emergency contacts in a dedicated callout box at the front of your journal.

Clinical partnership. Coordinate with healthcare providers for detox protocols, medication-assisted treatment, and co-occurring mental health support. Share your trigger ledger and replay summaries so clinicians can tailor care, then document their recommendations in your weekly review memo.

Relapse response map. Design an "if/then" playbook: if you take a first drink, then (1) pause intake, (2) notify support partner, (3) revisit belief audit, (4) book clinical follow-up. Store the plan physically (wallet card) and digitally (phone screenshot) so it is reachable during stress.

Crisis resources. List hotlines (local emergency number, SAMHSA helpline, regional crisis teams) and peer supporters.[3] Treat this list like a fire extinguisher—visible, checked monthly, and never neglected.

Legal and workplace considerations. Document any obligations (e.g., driving when sober, disclosure policies) and embed reminders in your action plan. Being proactive prevents cascading consequences that could trigger shame-based relapse.

9 Action Plan Roadmap

With the system components in place, convert them into a phased roadmap: detox (days 1–7), stabilization (weeks 2–6), and expansion (month 2 onward). Each phase gets a mission objective, key metrics, and review cadence.

[3] For readers in the United States, the SAMHSA National Helpline is available 24/7 at 1-800-662-HELP. Substitute the appropriate crisis lines for your country and keep them accessible in both digital and printed form.

Detox. Medical clearance, acute withdrawal management, hydration, sleep restoration, and initial support network activation. Track vital signs, cravings, and adherence to medical instructions.

Stabilization. Establish daily routines, populate the replacement portfolio, and run weekly replay sessions. Monitor metrics such as craving frequency, mood scores, therapy attendance, and financial savings.

Expansion. Layer in long-term goals—career development, creative projects, community leadership. Measure energy balance, relationship quality, contribution to others, and continued abstinence milestones.

Figure 6 sketches a dashboard concept that keeps these metrics visible. Use it as inspiration to build printable trackers or digital dashboards that fit your tooling.

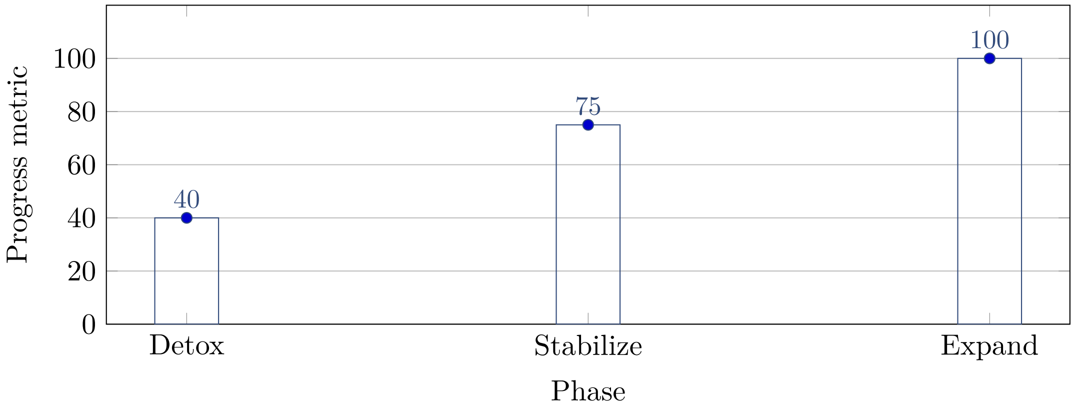

Figure 6: Example dashboard summarising completion of key sobriety milestones across detox, stabilization, and expansion phases. Each bar can be split by specific metrics such as sleep, cravings, and support engagement.

10 Upwards and Onwards!

Sobriety is not a void—it is altitude. Each trigger logged, script rehearsed, and policy refined lifts you from survival flying to purposeful navigation. The Deep Q-Life framework gives you cockpit instruments: states mark coordinates, actions chart flight paths, rewards confirm heading. With medical allies, accountability pods, and a purpose-rich schedule, you are no longer guessing; you are piloting.

Expect turbulence. Some weeks will shake the airframe, others will feel like cruising above the clouds. The work is to stay curious, replay the black box, and file a fresher flight plan. Celebrate every clean landing—mornings filled with clarity, weekends reclaimed, relationships restored. Those wins are the new rewards feeding your value function.

Keep iterating. Teach these tools to others, contribute data back to your support network, and expand the policy playbook for the next person deciding to quit. Upwards and onwards: the horizon is wide, and the longer you fly sober, the more destinations appear on your map.

Figures 7 and 8 summarize two practical decision heuristics that recur throughout the guide: widening the action set with structured awareness, and evaluating choices by total reward rather than short-term relief.

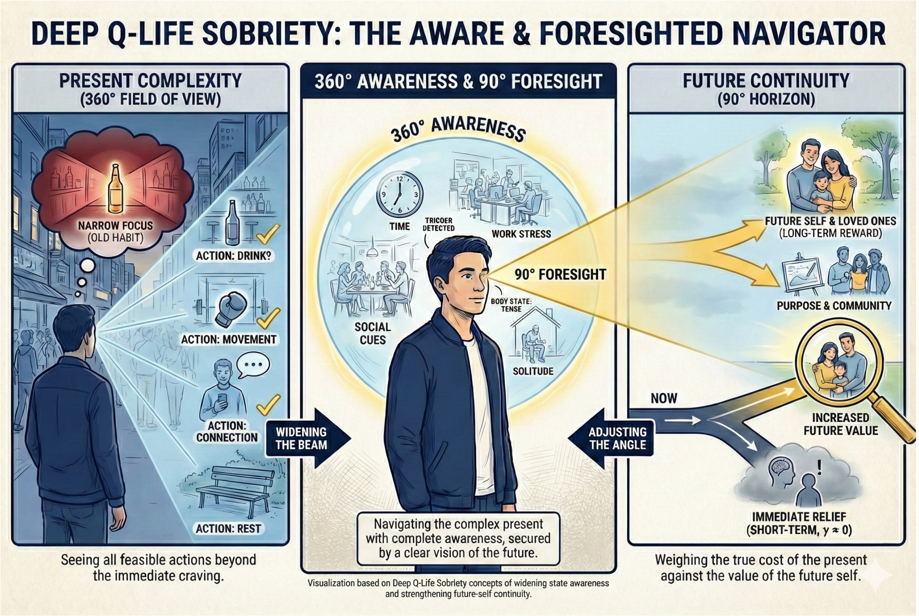
DEEP Q-LIFE SOBRIETY: THE AWARE & FORESIGHTED NAVIGATOR
PRESENT COMPLEXITY (360° FIELD OF VIEW)
NARROW FOCUS (OLD HABIT)
ACTION: DRINK?
ACTION: MOVEMENT
ACTION: CONNECTION
ACTION: REST
WIDENING THE BEAM
Seeing all feasible actions beyond the immediate craving.
360° AWARENESS & 90° FORESIGHT
360° AWARENESS
TIME
TRICOER DETECTED
WORK STRESS
90° FORESIGHT
SOCIAL CUES
BODY STATE: TENSE
SOLITUDE
Navigating the complex present with complete awareness, secured by a clear vision of the future.
Visualization based on Deep Q-Life Sobriety concepts of widening state awareness and strengthening future-self continuity.
ADJUSTING THE ANGLE
FUTURE CONTINUITY (90° HORIZON)
FUTURE SELF & LOVED ONES (LONG-TERM REWARD)
PURPOSE & COMMUNITY
NOW
INCREASED FUTURE VALUE
IMMEDIATE RELIEF (SHORT-TERM, γ ≈ 0)
Weighing the true cost of the present against the value of the future self.

Figure 7: Key summary: expand awareness from the immediate cue to a full 360-degree scan, then add 90 degrees of foresight toward likely next states and consequences.

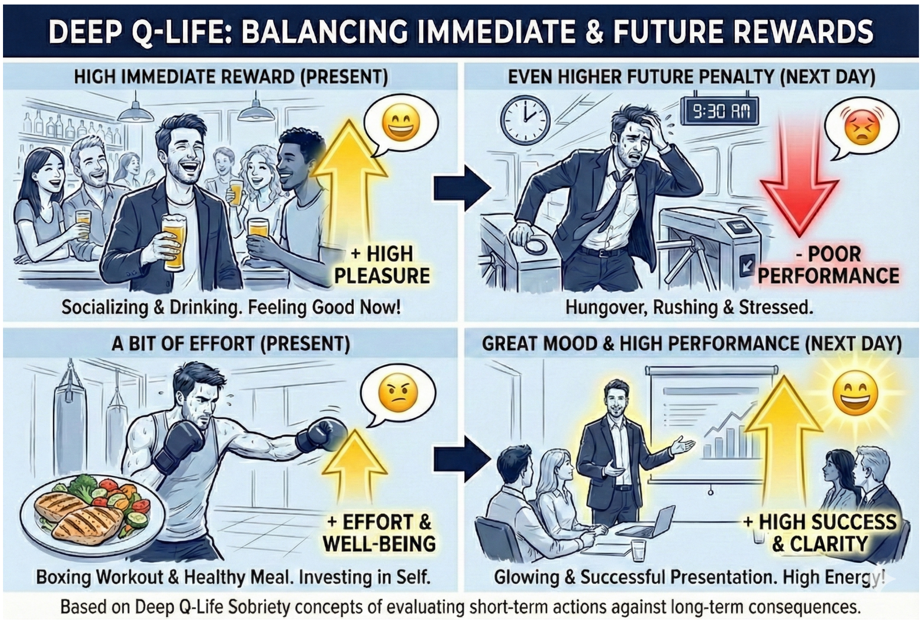
DEEP Q-LIFE: BALANCING IMMEDIATE & FUTURE REWARDS
HIGH IMMEDIATE REWARD (PRESENT)
+ HIGH PLEASURE
Socializing & Drinking. Feeling Good Now!
EVEN HIGHER FUTURE PENALTY (NEXT DAY)
9:30 AM
- POOR PERFORMANCE
Hungover, Rushing & Stressed.
A BIT OF EFFORT (PRESENT)
+ EFFORT & WELL-BEING
Boxing Workout & Healthy Meal. Investing in Self.
GREAT MOOD & HIGH PERFORMANCE (NEXT DAY)
+ HIGH SUCCESS & CLARITY
Glowing & Successful Presentation. High Energy!
Based on Deep Q-Life Sobriety concepts of evaluating short-term actions against long-term consequences.

Figure 8: Key summary: balancing reward tradeoffs between immediate fun (tonight), delayed costs (tomorrow's hangover), and higher-value alternatives (e.g., training, good food, and peak condition for next-day performance).

A Mindfulness and Future-Self Protocols

This appendix collects short “protocol cards” that expand the callouts in the main text. Each protocol is designed to be practical, time-bounded, and directly mapped to Deep Q-Life concepts such as state representation, action selection, and discounting.

Protocol 1: Mindful Breathing for Cravings (3 minutes)

When to use: in any state where craving intensity rises above baseline, especially before entering a high-risk environment.

Steps:

1. Sit or stand with feet on the floor. Gently lengthen your spine and relax your shoulders.
2. Inhale through the nose for a count of four, hold for one, exhale through the mouth for a count of six. Repeat this cycle for ten breaths.
3. With each exhale, silently label the moment: “noticing craving, choosing clarity”.
4. After ten breaths, quickly name your current state (location, emotion, body sensation) and write down one non-drinking action you will take next.

In RL terms, this protocol stabilises the observation of s and prevents impulsive, noise-driven action selection.

Protocol 2: Urge Surfing Body Scan (5 minutes)

When to use: when an urge feels overwhelming and you are tempted to “make it go away” quickly.

Steps:

1. Set a five-minute timer. Commit to no alcohol or other numbing actions until the timer ends.
2. Starting at your feet and moving upward, scan for where the urge is most vivid (tightness, heat, buzzing).
3. When you find a hotspot, place a hand there if possible and track the sensations moment by moment, noting changes in intensity, shape, or temperature.
4. Silently repeat for the duration: “This is a wave. Waves rise and fall. I can ride this without acting.”
5. When the timer ends, record a one-line observation in your journal (e.g., “Craving peaked at 8/10, dropped to 3/10 without a drink.”).

This trains the “observe and ride it out” action referenced in the main text, increasing the likelihood that it is selected instead of the old drinking response.

Protocol 3: Future-Self Visualization (5 minutes)

When to use: before evenings, social events, or travel days that historically lead to drinking.

Steps:

1. Close your eyes and imagine waking up tomorrow morning having stayed sober. Notice details: light, body sensations, calendar, interactions.
2. Extend the scene to three months from now, assuming continued sobriety. Visualise one concrete benefit (e.g., improved lab results, a completed project, a calmer home).

3. Extend again to one year of alcohol-free living. Picture a specific achievement or relationship that sobriety made possible.

4. For each time horizon, ask, “What action in the next two hours most increases this version’s quality of life?” Capture the answers in a few words.

5. Choose the replacement action that best serves all three future selves and commit it to your plan for the next block of time.

This protocol increases subjective future-self continuity and effectively nudges the internal discount factor γ upward.

Protocol 4: Morning Policy Preview (2–3 minutes)

When to use: at the start of the day, ideally after waking or during a morning routine.

Steps:

1. Review yesterday’s replay notes or a brief summary of the previous week.

2. Ask, “What is the single soft factor I want to highlight today?” Choose one: state-awareness, urge surfing, self-compassion, or future-self focus.

3. Write a single sentence policy for today, such as “When I notice tension on the commute, I will run the Three-Breath State Check before deciding what to do next.”

4. Optionally, place a visual cue (sticky note, lock-screen reminder) that echoes this sentence.

Viewed through Deep Q-Life, this is a lightweight, daily policy update that makes one action rule salient and testable.

Protocol 5: Evening Replay and Compassion Log (3–5 minutes)

When to use: at the end of the day, before bed or during wind-down.

Steps:

1. Note one salient craving event from the day and record it as (s, a, r, s') in your journal using your chosen shorthand.

2. Add one brief line for each component: what you felt, what you did, what immediate reward you noticed, and one delayed consequence.

3. If there was a slip or a close call, write a compassionate debrief line of the form, “Mission log: [time], [trigger], [action]. Next time I will adjust [state, action, or support] by [specific tweak].”

4. Close with a gratitude or acknowledgement phrase, such as “I am learning” or “Data collected; policy improving.”

This practice keeps the learning loop active, prevents all-or-nothing thinking, and feeds higher-quality data into weekly target network updates.

References

[1] Y. J. Hilpisch. *Deep Q-Life: Markov, Bellman, and Habits.* 2025. Available at `https://hilpisch.com/dql_habits.pdf`.

[2] R. Bellman. *Dynamic Programming.* Princeton University Press, 1957.

[3] R. S. Sutton and A. G. Barto. *Reinforcement Learning: An Introduction.* 2nd ed., MIT Press, 2018.

[4] V. Mnih, K. Kavukcuoglu, D. Silver, et al. Human-level control through deep reinforcement learning. *Nature* 518, 2015.

[5] J. Clear. *Atomic Habits.* Avery, 2018.

[6] W. Wood. *Good Habits, Bad Habits.* Farrar, Straus and Giroux, 2019.

Disclaimer. This document is for educational purposes only and does not constitute medical, psychiatric, or legal advice. Quitting or reducing alcohol can carry significant health risks; readers should always consult qualified healthcare professionals and follow local regulations when making changes to their consumption, medications, or treatment plans.