Available online at www.sciencedirect.com

ScienceDirect

Current Opinion in Behavioral Sciences

Neural autopilot and context-sensitivity of habits

Colin F Camerer[1,2] and Xiaomin Li[1]

This paper is about the background of two new ideas from neuroeconomics for understanding habits. The main idea is a two-process 'neural autopilot' model. This model hypothesizes that contextually cued habits occur when the reward from the habitual behavior is numerically reliable (as in related models with an 'arbitrator'). This computational model is lightly parameterized, has the essential ingredients established in animal learning and cognitive neuroscience, and is simple enough to make nonobvious predictions. An interesting set of predictions is about how consumers react to different kinds of changes in prices and qualities of goods ('elasticities'). Elasticity analysis expands the habit marker of insensitivity to reward devaluation, and other types of sensitivities. The second idea is to use machine learning to discover which contextual variables seem to cue habits, in field data.

Addresses
[1] Division of Humanities and Social Science, California Institute of Technology, Pasadena, CA, United States
[2] Computational and Neural Systems, California Institute of Technology, Pasadena, CA, United States

Corresponding author: Camerer, Colin F (camerer@hss.caltech.edu)

Current Opinion in Behavioral Sciences 2021, **41**:185–190

This review comes from a themed issue on **Value-based decision-making**

Edited by **Laura Bradfield** and **Bernard Balleine**

For a complete overview see the Issue and the Editorial

Available online 10th September 2021

https://doi.org/10.1016/j.cobeha.2021.07.002

Defining habits

For such an apparently simple behavior, habit is tricky to define [1]. Daw and Rünger [2, p. 291] suggest that 'the common elements are habit learning through repeated responding so as to form context-response associations in memory, and automated habit performance that is relatively insensitive to changes in the value or contingency of response outcomes'.

In this paper, we describe two neuroeconomic computational approaches designed to bridge controlled lab experiments with messier elements of everyday behavior. We use a concept of predictability as a hallmark of context-sensitivity in habits.

Beyond history-dependence: neural autopilot

In general, neuroeconomics seeks to combine the best methods, and data from both economics and cognitive neuroscience to create theories that are mathematically insightful, useful across biological and social sciences, and also correspond to actual neural mechanisms.

In the case of habit formation, this combination is challenging. Outside of psychology and neuroscience, habits are defined by a hypothesized relation between past choices which increases current preference (subjective value) [3] (cf. [4,5]). If this is true and causal, then artificially inducing people to engage in an activity (e.g. paying to encourage them go to the gym) should increase the activity. However, the evidence for this causal link is mixed. Post-treatment activity usually drops off rapidly (e.g. [6]).

The neuroeconomic approach hypothesizes two types of choice control systems in the brain — model-free and model-based [7]. Habit, cued associatively by context and surprisingly insensitive to reward changes, is an 'overtrained' model-free system. The goal-directed model-based system, in contrast, always keeps goals in mind and approximates ideal strategies.

A few proposals have suggested how the brain combines these two systems. In Keramati *et al.* [8], a normative model dynamically integrated both habitual and goal-directed processes based on speed/accuracy trade-offs and expected information value. An agent decides to deliberately calculate the value of an action (i.e. use the model-based system) only if the potential benefit of this action, $VPI(s,a)$, exceeds a time-sensitive cost. Otherwise, action values are recalled from past experiences (that is the habit). Another hybrid model [9] plans through decision trees to a limited depth, then substitutes habit-based cached action-state $Q(a,s)$ values (much like chess algorithms do).

Daw *et al.* [10] suggested a neural 'arbitration' between model-free (habitual) and model-based control, based on uncertainty, for which there is now fMRI evidence in Lee *et al.* [11•].

A neural autopilot theory introduced in [12] uses similar elements about choice values r from reward prediction error (RPE), and learning about 'reward reliability' (unsigned RPE). In neural autopilot, consumer i has a running valuation of product j at time t, denoted r_{ijt}, which is a weighted average of past r_{ijt} and reward prediction error. The choice at time t is $c_i(t)$. Reward prediction is updated according to r_{ijt}:

$$r_{ijt} = \begin{cases} r_{it-1} + \rho(u_{ijt-1} - r_{it-1}), \text{if } c_i(t-1) = j \\ r_{ijt-1}, \text{otherwise} \end{cases}$$

Note that contextual cues should enter here if the current choice depends on context variables such as which location they are in, time of day, internal hunger or craving states, etc. We do not use notation to denote such context-dependence but is easy to expand the notation to include such effects. Reward *reliability* d_{ijt} is driven by the absolute value of (often called 'unsigned') prediction error. It is updated according to:

$$d_{ijt} = \begin{cases} (1-\lambda)d_{ijt-1} + |u_{it} - r_{ijt}|, \text{if } y_{i,t-1} = j \\ (1-\lambda)d_{ijt-1} + \alpha, \text{otherwise} \end{cases}$$

When deciding whether to choose the product j, or a different product in the choice set, the consumer will not optimize. Instead, she chooses between two decision rules. She starts by recalling her last choice and checking if its reward reliability is sufficiently low. In this specification, reliability is measured by unsigned prediction error.[3] If the reliability is below a threshold θ, she repeats her choice from the last period $y_{i,t-1}$. Otherwise she maximizes predicted reward (in model-based mode).

$$y_{i,t} = \begin{cases} y_{i,t-1}, \text{if } d_{iy_{i,t-1}t} < \theta \\ \text{argmax} r_{ijt}, \textit{otherwise} \end{cases}$$

It is well-known that for most goods and services, short-run price 'elasticities' (sensitivity of percentage choice to percentage price) are much smaller than long-run elasticities, and are often close to zero. This property can be derived from a neural autopilot habit model; even better, parameters of the model predict the time of at which short-run transitions to the long-run happen (which is an unsolved problem in economics). For certain data sets, the neural autopilot theory can be tested empirically by estimating free parameters.

One nonobvious prediction of the neural autopilot model is that when making habitized choices, a person will not be tracking prediction errors of unchosen outcomes, and will therefore not know if there is an unchosen choice that has become more rewarding. (Adding some stochasticity or directed exploration will change this property.) A neat illustration comes from a 48-hour strike at the London Tube (subway) [13•]. Some Tube commuters were forced to find a different route to work because their regular trains were not operating. About 5% of the people who found a better route then switched to the better route, saving an estimated £138.

A major challenge for any neural autopilot approach is to relate reward reliability to the mountain of evidence from animal learning. Behavior rewarded on random interval training schedules (rewards based on time) appears to be more prone to habitization than random ratio schedule training (rewards based on actions) [14,15]. The explanations for the difference in habitization is that interval reward either weakens the correlation between reward rate and action rate, or creates incidental association with action timing [16]. An important open question is whether stronger habituation from interval rewards can somehow be reconciled with more reliable reward from interval training (that is required to make the differential habitization fit the neural autopilot model). One possibility is endogeneous behavior by animals– if they have learned to withhold action just after an interval reward (crudely anticipating the substantial time interval before the next reward) and act just before a pending interval reward, that behavior will increase reward reliability. Another possibility is that reward rates are somehow averaged across trials rather than linked to single-trial actions (as is assumed in the current neural autopilot specification). Note also that fixing a habit threshold σ, different rates of reward will generate different reliabilities, which predicts that different degrees of habitization will be associated with different reward rates (in either interval or ratio schedules).

In Milkman *et al.* [17], a field experiment incentivizing gym attendance was reported which provides indirect evidence of potential importance of reward reliability. During a four week intervention, Google workers were paid \$3 or \$7 to go to a workplace gym for 30 m, either during a common 2-hour window of their initial choosing, or at any flexible time. Behavior in the post-intervention period indicates whether persistent behavioral habits were formed. As is typical in such studies, the intervention worked when it was active (increasing visits by 0.64/week) but had a smaller post-intervention effect in the next 4 weeks (+0.20 extra visits). Removing the larger \$7 incentive apparently generated a larger jump in reward *un*reliability and reduced visits by more than removing the \$3 incentive (−0.56/week versus −0.34/week). Furthermore, incentivized flexible attendance increased post-intervention visits more than fixing attendance in the 2-hour window. This is consistent with the hypothesis that flexible busy workers were able to choose more reliably-rewarding times to go to the gym, while the routine-bound workers had more variability in

[3] Other measures of reward reliability have been used. Lee *et al.* [11•] use a categorized fraction of time that reward prediction error is zero. Finding the empirically appropriate measure of reward reliability, and its neural encoding, is an important solveable question for future research.

subjective value and also more variability in reward prediction errors.

Harris and Kessler [18] explored habits using data from use of stationary exercise equipment. They use rainfall as a quasi-experimental variable hypothesized to reduce current exercise, and to causally reduce future exercise (if current exercise creates habit). They identify such an effect of initial exercise which is large in magnitude (though not precisely estimated): One extra workout in a four-week initial period leads to two extra workouts in the next six weeks. Their analyses further suggest that successful interventions need to create daily workouts to have substantial post-intervention success (p. 705).

Cues and contexts

Many human studies have shown the sensitivity of apparent habits to contextual cues. Laboratory results are tentatively convincing. The next step is to understand similar effects in field data outside the lab.

Two particularly clever lab experiments illustrate context effects. In Neal *et al*. [19], when more (self-reportedly) habitized people ate popcorn in an actual movie theater (compared to a control), they ate as much stale popcorn as fresh popcorn (less-habitized subjects at less stale popcorn). Being in the theater is the context variable. Another neat experiment recorded subjects' speech intensity (DB) after priming by searching for kitchen images (control) or stadium images (treatment). Those who visit stadiums more often were more likely to speak louder when the stadium context image was present [20].

Using a large set of data on recorded health care workers' handwashing [21] found that washing habits were interrupted by context changes — longer breaks from work and working in unfamiliar hospital locations.

In the clinical literature on addiction, context-sensitivity is called 'cue reactivity'. In typical experimental paradigms, addicts are exposed to sensory cues which have been previously Pavlovian-associated with drug use. The goal is to find which cues create biological craving and drug use.

Meta-analysis shows that cannabis cues (such as pictures, videos, handling a joint) cause craving, whether measured by self-report or by psychophysiology (EEG, EDA, heart rate) around $d = 0.60$ [22]). A robust type of internal cue is a negative effect, induced experimentally — by social stress, rumination, shock anticipation, etc. Meta-analysis has shown that negative affect induction has an effect around $d = 0.30$–0.40 on both craving and alcohol use [23], and on tobacco craving [24]. Self-reported craving states correlate around $d = 0.30$ with 'nonautomatic' tobacco use (planning where to buy cigarettes) and $d = 0.15$ with 'automatic' use (e.g. lighting up), though there is an apparent bias to overpublish positive effects [25].

An early illustrative example of testing for different context variables in the field used ecological momentary assessment (EMA) [26]. People were asked via an electronic diary to report what they were currently doing, and their craving, drug use, and emotions. The strongest cue associated with craving and drug use was when addicts were in the presence of the kinds of cash bills ($10 or more), which were often used to buy drugs. In Kirchner *et al*. [27•], both EMA and the geospatial locations of 475 smokers who were trying to quit were tracked using phone GPS. People were more likely to lapse and smoke when there were physically near point-of-sale tobacco locations. The effect is stronger when the craving level is low.

In Buyalskaya *et al*. [28], a novel approach was taken to studying context-sensitivity using large panel data sets, in which many people were observed repeatedly over long several-month spans. To find cues associated with behavior, they consider a large number of candidate context predictors. Machine learning (LASSO) is then used to select predictors and avoid overfitting.

Their procedure is called 'predicting context sensitivity' (PCS). It was applied to two data sets, on gym attendance and how often hospital workers washed their hands using automated sanitizers in patients' hospital rooms (see Box 1). The predictive accuracy of a logistic model of 0–1 behavior (either attending the gym, or washing hands), was measured conventionally by the area under a ROC curve (called AUC). The individual-level values were around AUC = 0.65–0.85 (where 0.50 is random and 1.0 is perfect).

The authors also tested whether hospital workers were more insensitive to the likely change in reward value from handwashing in the last room of the day (assuming reward value falls because their hands do not need to be clean to protect patients any longer). This test is a direct analogue to the method from animal learning but without lab control. They did not find an effect of habit on reward insensitivity. However, more tests with a better reward-change proxy are worthwhile.

The PCS procedure can also be used to see if the context-sensitivity is becoming more predictable over time, as a habit is being formed (statistically, is AUC increasing?). Statistical methods can then be used to estimate how long it takes habits to form based on how AUC increases over time (see Box 2). While many people do not seem to form stable habits of this type (i.e. their predictability does not increase over time), for the 30–45% who do so the estimated time to habit formation is around 200 days for gym attendance and 14 daily shifts for hospital workers.

Box 1 Predicting context-sensitivity (PCS)

Buyalskaya *et al*. [28], they used LASSO to predict context-sensitive handwashing behaviors in a hospital. They leverage a dataset collected via a RFID technology that monitors whether an individual washes their hands in each hospital room they enter. A total of 5246 healthcare workers across 30 different hospitals, were tracked over about a year (40M data points).

Each data point is accompanied by a timestamp, as well as the room and hospital where two opportunities to wash arose, when both entering and exiting. Candidate context variables include the time of day, time spent working, previous room and shift handwashing compliance, and indicators for entry or exit. (In general, the procedure allows *any* context variables to be included.)

LASSO regression is used for 'selecting' predictive variables and 'shrinking' variables which are not predictive to zero (to avoid overfitting). For each person, a LASSO regression establishes a single measure of predictability across the entire sample — the area under the ROC curve (AUC). Higher AUC means the behavior is more predictable. AUCs were around 0.75–0.85.

Across individuals, AUC is not correlated with the frequency with which the behavior is performed ($r = -0.06$). In other words, whether people wash their hands frequently or not, in general, is *not* associated with how contextually predictable they are. This is an important fact because habits are sometimes said to require frequent behavior for habitization to take place.

However, some infrequent behaviors can be context-sensitive and habit-like. For example, Aldrich *et al*. [29] suggest that voting can be habit-like, in the specific sense that a causal nudge to vote once influences future voting, but only if the context is held fixed (where 'context' is living in the same home).

Measures of predictability and the set of important feature predictors can be used in several ways: How high is predictability? Can less-predictable people change habits more easily? What variables are most predictive?

Table 1 shows the strongest predictors of handwashing at the aggregate level. The most important context variable is handwashing compliance during their last shift. A room entry indicator is negative for 77% of workers, which means workers are more likely to wash when exiting than when entering (this is bad if it reflects workers protecting themselves from patient germs rather than the other way around). Time since the start of a shift ('Time at work') is a negative predictor of handwashing for 42%. Room compliance of others is generally a positive predictor (for 66%). This likely reflects the possibility that rooms with sicker patients, where sanitizers are especially salient, or where workers notice their peers washing, have higher overall washing rates. Many variables are highly *un*important (mostly zero LASSO coefficients) (not shown), including times of day, months, and time off before the current shift.

Conclusion

Decades of careful research on animal learning, and a smaller body of recent work on human learning, have shown some basic principles about how habit seems to work in small-scale tightly controlled settings. Since so many important questions about high-consequence human behavior hinge around shedding bad everyday habits and acquiring good ones, linking the mechanistic theories to observable everyday behavior in large data sets is now feasible. This opinion described two ways to make that link by developing and testing a two-system neural autopilot theory and using machine learning to identify individual context-sensitivity.

One weakness of the PCS analysis is that there is no information about the automaticity of habit response. It is likely that some parts of the action chain in behavior like gym attendance initiation are explicit and deliberate (e.g. deciding when to go during a day) and other elements are automatic (e.g. finding and opening a familiar locker). Better data will help tell the difference.

Table 1

Context predictors of hospital hand washing

	Importance	Q1	Median	Q3	% zero	% positive	% negative	Homog index
Compliance last shift	0.77	0.66	0.70	0.92	0	100	0	96
Entry indicator	0.35	−0.33	−0.28	−0.04	18	5	77	68
Compliance last opp. × Entry indicator	0.13	0.00	0.00	0.21	49	47	4	39
Compliance last opp. × Time since last opp	0.12	0.00	0.00	0.00	54	1	45	40
Compliance within episode	0.12	0.00	0.01	0.14	33	51	16	31
Time since last opp.	0.09	0.00	0.00	0.00	61	24	15	5
(Time since last opp.)^2	0.08	0.00	0.00	0.00	74	7	18	7
Room compliance of others	0.08	0.04	0.05	0.12	32	66	2	60
Time at work	0.08	0.00	0.00	0.00	54	4	42	34
Compliance last opp. × (Time since last opp)^2	0.07	0.00	0.00	0.00	74	20	5	11
Prev. room compliance	0.07	0.03	0.04	0.11	32	65	2	59

Statistics of feature importance. Importance is the absolute value of standardized LASSO coefficients averaged across individuals. Q1, Median, and Q3 are the coefficient values for first (lowest), second, and third quantiles of the sample. % zero, % positive, and % negative capture the percentage of the individual LASSO models which had coefficients that had zero, positive, and negative values, respectively. The Homogeneity Index is the absolute value of the difference between % positive and % negative.

Box 2 Estimating when habits form

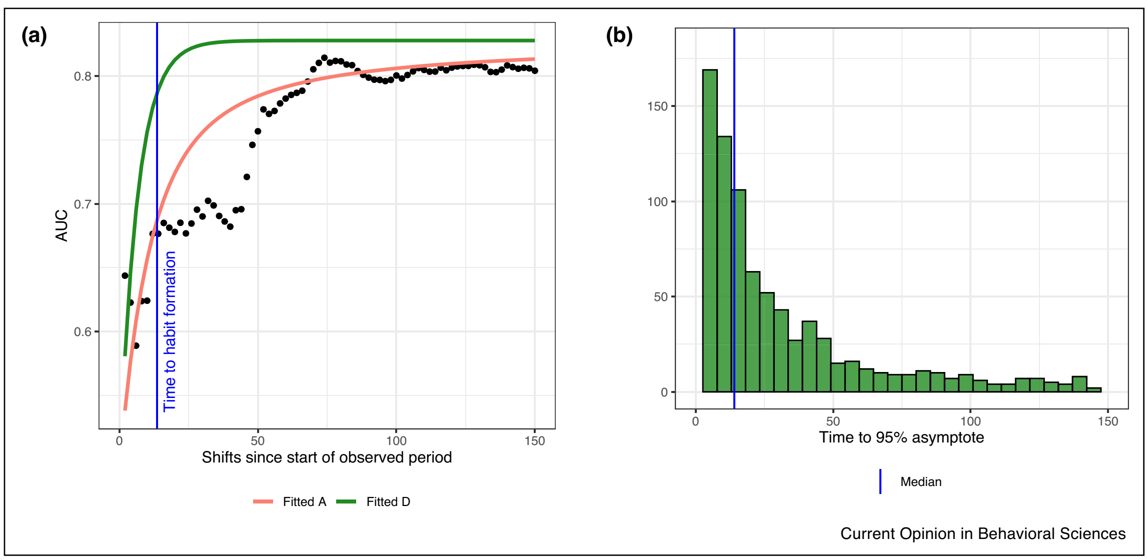
How quickly habits form is an important question because it sheds light on neurophysiological mechanisms and is also of practical importance (e.g. to promote or change habits). In Buyalskaya *et al.* [28], machine learning was applied to measure habit formation in terms of statistical predictability of behavior from context variables.

The method assumes there is an instantaneous predictability function $D_i(t)$ which can be approximated by the function $a - b^{-ct}$ following Lally *et al.* [30•]. Because the AUC requires a substantial sample to estimate, $D_i(t)$ is derived from a related function $A_i(t)$ given by

$$A_i(t) = \frac{1}{t}\int_0^t D_i(s)ds$$

Calculus shows that $A_i(t) = a_i - \frac{b_i[1-\exp(-c_i t)]}{c_i t}$. Nonlinear least squares is used to fit the empirical $A_i(t)$ to each individual i's AUC sequence and obtain the estimates $\hat{a}_i, \hat{b}_i, \hat{c}_i$. The time to habit formation for person i is defined the time it takes for $D_i(t)$ to reach 95% of its estimated asymptote a_i (which is $T_i^* = -\ln(a_i/20b_i)/c_i$). This function is plotted in green in Figure 1a. Figure 1b is a histogram of T_i^* values across all workers.

Figure 1



Development of habit formation.

Conflict of interest statement

Nothing declared.

References and recommended reading

Papers of particular interest, published within the period of review, have been highlighted as:

- • of special interest

1. Balleine BW, Dezfouli A: **Hierarchical action control: Adaptive collaboration between actions and habits**. *Front Psychol* 2019, **10**:2735.

2. Wood W, Rünger D: **Psychology of habit**. *Annu Rev Psychol* 2016, **67**:289-314.

3. Becker GS, Murphy KM: **A theory of rational addiction**. *J Polit Econ* 1988, **96**:675-700.

4. Laibson D: **A cue-theory of consumption**. *Quart J Econ* 2001, **116**:81-119.

5. Bernheim B, Rangel A: **Addiction and cue-triggered decision processes**. *Am Econ Rev* 2004, **94**:1558-1590.

6. Milkman KL, Patel MS, Gandhi L, Graci H, Gromet D, Ho H, Kay J, Lee T, Akinola M, Beshears J *et al.*: **A mega-study of text-based nudges encouraging patients to get vaccinated at an upcoming doctor's appointment**. *Nature* 2021. (in press).

7. Poldrack RA: *Hard to Break: Why Our Brains Make Habits Stick*. Princeton University Press; 2021.

8. Keramati M, Dezfouli A, Piray P: **Speed/accuracy trade-off between the habitual and the goal-directed processes**. *PLoS Comput Biol* 2011, **7**:e1002055.

9. Keramati M, Smittenaar P, Dolan RJ, Dayan P: **Adaptive integration of habits into depth-limited planning defines a habitual-goal-directed spectrum**. *Proc Natl Acad Sci U S A* 2016, **113**:12868-12873.

10. Daw ND, Niv Y, Dayan P: **Uncertainty-based competition between prefrontal and dorsolateral striatal systems for behavioral control**. *Nat Neurosci* 2005, **8**:1704-1711.

11. • Lee S, Shimojo S, O'Doherty J: **Neural computations underlying arbitration between model-based and model-free learning**. *Neuron* 2014, **81**:687-699

This article creates a computational model of reliability of model-free (MF) and model-based (MB) systems and shows neural circuitry in which inferior LPFC acts like an arbitrator that encodes reliability and psychophysiologically modulates MF and MB signals.

12. Camerer C, Landry P, Webb R: *The Neuroeconomics of Habit*. 2018. Available at SSRN 3752193.

13. • Larcom S, Rauch F, Willems T: **The benefits of forced experimentation: striking evidence from the London underground network**. *Quart J Econ* 2017, **132**:2019-2055

This article indicates that people stay on suboptimal commuting choices on the London underground, due to informational imperfections, but some of them change after forced experimentation due to a labor strike.

14. Dickinson A, Nicholas DJ, Adams CD: **The effect of the instrumental training contingency on susceptibility to reinforcer devaluation**. *Quart J Exp Psychol* 1983, **35**:35-51.

15. Perez OD, Dickinson A: **A theory of actions and habits: the interaction of rate correlation and contiguity systems in free-operant behavior**. *Psychol Rev* 2020, **127**:945-971.

16. Killeen PR: **Mathematical principles of reinforcement**. *Behav Brain Sci* 1994, **17**:105-135.

17. Beshears J, Lee HN, Milkman KL, Mislavsky R, Wisdom J: **Creating exercise habits using incentives: The trade-off between flexibility and routinization**. *Manag Sci* 2021, **67**:4139-4171.

18. Harris MC, Kessler LM: **Habit formation and activity persistence: evidence from gym equipment**. *J Econ Behav Organ* 2019, **166**:688-708.

19. Neal DT, Wood W, Wu M, Kurlander D: **The pull of the past: when do habits persist despite conflict with motives?** *Pers Soc Psychol Bull* 2011, **37**:1428-1437.

20. Neal DT, Wood W, Labrecque JS, Lally P: **How do habits guide behavior? Perceived and actual triggers of habits in daily life**. *J Exp Soc Psychol* 2012, **48**:492-498.

21. Wellsjo AS: *Simple Actions, Complex Habits: Lessons From Hospital Hand Hygiene*. 2021. Working Paper.

22. Norberg MM, Kavanagh DJ, Olivier J, Lyras S: **Craving cannabis: a meta-analysis of self-report and psychophysiological cue-reactivity studies**. *Addiction* 2016, **111**:1923-1934.

23. Bresin K, Mekawi Y, Verona E: **The effect of laboratory manipulations of negative affect on alcohol craving and use: a meta-analysis**. *Psychol Addict Behav* 2018, **32**:617.

24. Heckman BW, Kovacs MA, Marquinez NS, Meltzer LR, Tsambarlis ME, Drobes DJ, Brandon TH: **Influence of affective manipulations on cigarette craving: a meta-analysis**. *Addiction* 2013, **108**:2068-2078.

25. Gass JC, Motschman CA, Tiffany ST: **The relationship between craving and tobacco use behavior in laboratory studies: a meta-analysis**. *Psychol Addict Behav* 2014, **28**:1162.

26. Epstein DH, Willner-Reid J, Vahabzadeh M, Mezghanni M, Lin J-L, Preston KL: **Real-time electronic diary reports of cue exposure and mood in the hours before cocaine and heroin craving and use**. *Arch Gen Psychiatry* 2009, **66**:88-94.

27. • Kirchner TR, Cantrell J, Anesetti-Rothermel A, Ganz O, Vallone DM, Abrams DB: **Geospatial exposure to point-of-sale tobacco: real-time craving and smoking-cessation outcomes**. *Am J Prev Med* 2013, **45**:379-385

This study finds a positive correlation between cues in field settings (exposure to point-of-sale tobacco markets measured by GPS via mobile phones) and smoking.

28. Buyalskaya A, Ho H, Li X, Duckworth A, Milkman K, Camerer C: *Predicting Context-Sensitivity and Habit Formation in Field Data Using Machine-Learning*. 2021. Submitted for publication.

29. Aldrich JH, Montgomery JM, Wood W: **Turnout as a habit**. *Polit Behav* 2011, **33**:535-563.

30. • Lally P, van Jaarsveld CH, Potts HW, Wardle J: **How are habits formed: modelling habit formation in the real world**. *Eur J Soc Psychol* 2010, **40**:998-1009

This study develops a novel dynamic approach using asymptotic regression over time to identify the length of time needed for forming habits, using self-reported habit strength.