Neural substrates underlying the expectation of rewards resulting from effortful exertion

Authors: Aram Kim[1,2], Vikram S. Chib[1,2,3]*

Affiliations:

[1] Department of Biomedical Engineering, Johns Hopkins School of Medicine; Baltimore, MD, USA.

[2] Center for Movement, Kennedy Krieger Institute; Baltimore, MD, USA.

[3] Kavli Neuroscience Discovery Institute, Johns Hopkins University; Baltimore, MD, USA.

*Corresponding author. Email: vchib@jhu.edu

Abstract: Why do people persist—or stop—when rewards require effort? Such choices are often framed as a trade-off between reward and cost, yet this conceptualization overlooks a fundamental determinant: expectations. Economic theories propose that labor supply is reference-dependent, responding to deviations from expected earnings, but whether such reference points govern effort and how they are implemented in the brain has remained unknown. Using an incentivized grip-force paradigm with functional magnetic resonance imaging, we experimentally imposed reward expectations independent of outcomes. Effort provision shifted systematically with these expectations, demonstrating that individuals evaluate exertion relative to a reference point rather than absolute incentives. Activity in the ventral striatum encoded both prospective reward expectations and deviations between realized and expected earnings, and these signals predicted inter-individual differences in reference-dependent effort. These findings identify a neural mechanism by which expectations recalibrate the value of exertion, revealing reference dependence as a fundamental principle of motivated behavior.

One-Sentence Summary: By demonstrating that reward expectations serve as neural reference points for effort valuation, we reveal reference dependence as a fundamental principle of motivated action.

Main Text:

Introduction

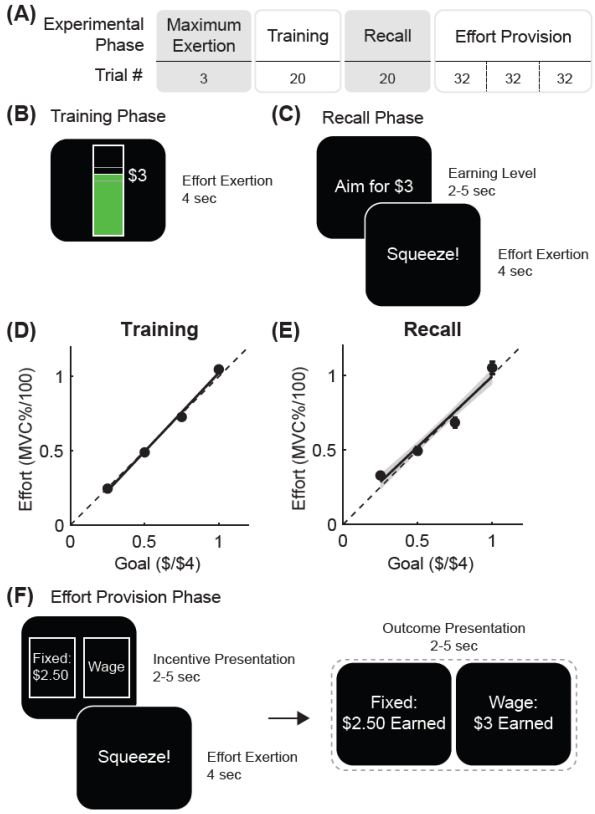
Expectations critically shape decisions to engage in effortful activity. For example, a rideshare driver might decide how many riders to pick up based on their earnings expectations, balancing the perceived effort of driving with the expected compensation. Indeed, many field experiments have consistently observed that reward expectations shape willingness to participate in effortful activities (*1*–*5*). For instance, taxi drivers determine how long they work based on their history of hours worked and their expected daily income (*1*), while professional athletes play more vigorously when they expect to win a game based on prior experience (*2*). These real-world examples underscore the role of expectations in shaping decisions about effortful activities. In this view, reward expectations serve as a reference point that motivates effortful exertion (*6*, *7*).

Despite the fundamental role of reference points in motivating human performance during effortful activity, little is known about how the brain encodes these expectations in the context of effort-based decision-making and exertion. A previous neuroimaging study of reference-dependent valuation in market transactions found that ventral striatal (VS) activity was associated with the computation of the relative value of economic goods with respect to an internally defined reference point (*8*). In value-based decision-making, VS activity has been shown to encode subjective utility (*9*), reward expectations (*10*), and reward prediction errors (*10*–*13*). Together, these studies suggest that VS could play a key role in representing reference-dependent preferences

that guide effortful exertion; however, the role of VS in reference-dependent effort preferences has not been directly tested.

We investigated the neural processes through which reference dependence influences the motivation to exert effort. We hypothesized that individuals' earnings expectations would serve as a reference point that influences their effort exertion – as an individual's expectation of reward increased, they would be more willing to exert effort. This hypothesis is based on previous field and laboratory experiments showing that people tend to work longer and earn more when they expect to earn more (*14–17*). We also hypothesized that reference-dependent effort provision would have its basis in signals in the VS such that activity in the VS would be modulated by the magnitude with which actual earnings deviated from expected earnings. This hypothesis is based on the role of the VS in encoding reward outcome expectations as the difference between expected and actual outcomes (i.e., reward prediction error) (*10–13*).

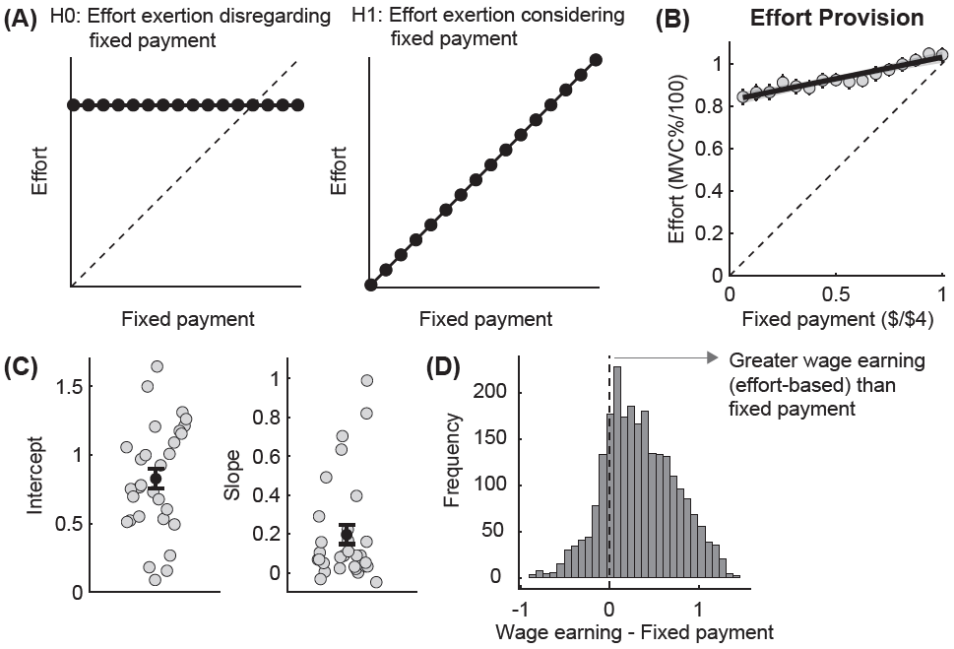
Results



To test these hypotheses, we asked participants to perform an incentivized physical effort task and used functional magnetic resonance imaging (fMRI) to examine neural activity in the VS during effort exertion and the experience of performance outcomes (Fig 1A). Participants exerted effort through grip force and began the experiment by producing their maximum grip force. This maximum voluntary contraction (MVC) was used to normalize each participant's strength for effort exertion. Participants then performed a Training Phase in which they learned to associate levels of effort with the amount of money earned (Fig 1B) – higher levels of exertion were associated with higher earnings, and maximum exertion corresponded to $4. To confirm that participants were adequately trained on the effort/earning associations, participants performed a Recall Phase to determine whether they could accurately exert effort to achieve a target level of earnings (Fig 1C). Participants successfully completed the

Training Phase (t(118) = 61.44, p < 0.001) (Fig 1D) and learned the effort/earning associations (t(118) = 22.73, p < 0.001) (Fig 1E).

Following the Training and Recall phases, participants were presented with the main Effort Provision Phase, in which options were either a payment proportional to the amount of effort exerted (Wage Rate Payment; as experienced during the Training Phase) or a fixed monetary payment, irrespective of effort exerted (Fixed Payment) (Fig 1F) (*14*). The outcome was realized at the end of each trial, and each option had an equal probability of occurring. On each trial of the Effort Provision Phase, participants had to decide how much effort to exert, given the fixed payment presented and the uncertainty about which option would be randomly selected. By observing the amount of effort a participant exerted in response to different fixed payment levels, we could infer how earnings expectations influence effort provision.



We hypothesized that participants could predominantly adopt two approaches when deciding how much effort to allocate, given the possibility of receiving either

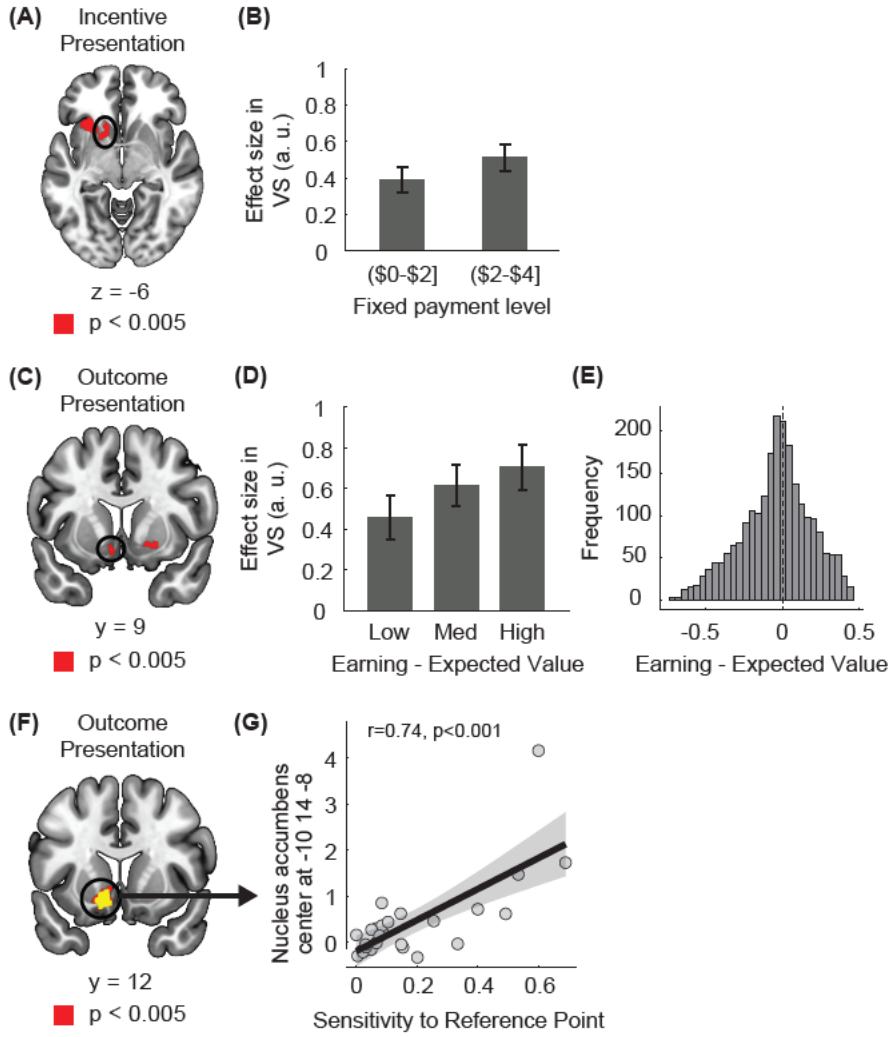
earnings proportional to their exertion or a fixed payment. In one case, if an individual felt that the effort was not costly or aversive, they could exert their maximum effort to maximize potential wage earnings and the expected outcome value of a trial, while disregarding the level of the fixed payment (Fig 2A, left panel). In another case, if participants find effort costly and aversive and want to avoid exerting effort that may not lead to a wage rate payment, they could modulate their exertion levels to account for the fixed payment amount in hopes of reducing their exertion to ensure that effort is not wasted (Fig 2A, right panel). How much effort is exerted to earn the wage rate, and an individual's willingness to overshoot the fixed payment could reflect the individual's cost of effort and effort-based incentive motivation.

In this context, if a participant's accumulated wage earnings are lower than the fixed payment, receiving the wage earnings falls short of their expectations, as they partly expected to get the (higher) fixed payment. Essentially, expectations provide an additional incentive for an individual to work hard, as they close the gap between possible outcomes and expectations. The incentive effect may be reversed once accumulated piece-rate earnings exceed the fixed payment, in which case the participant might end up with the (lower) fixed payment, reducing the incentive to exert effort. Reference dependence, therefore, shifts the optimal effort towards the level that equates accumulated piece rates with the fixed payment. Models that assume a reference point in expectations predict that increasing the size of the fixed payment will tend to increase overall effort, and that the propensity to stop exerting effort is especially high when piece-rate earnings equal the fixed payment (*14*, *18*).

Consistent with a strategy in which the fixed payment modulated exertion levels, we found that participants exerted more effort as the fixed payment increased ($t(1426) =$

13.70, $p < 0.001$, Fig 2B), and there was considerable inter-individual variability in effort across fixed payment values (Fig 2C). Notably, the majority of participants exhibited a positive intercept in the relationship between the fixed payment and effort (Fig 2C). Moreover, in most trials, the effort provided exceeded the effort required for the trial's reference point (Fig 2D). These results suggest that the cost of exertion did not exceed the value of the fixed payment. Consistent with previous behavioral studies of reference-dependent effort provision (1–5, 14, 15), our results suggest that the fixed payment influences participants' exogenous expectations and shapes motivation to exert effort.

To test how the brain encodes the reference point for effort provision at the time of incentive presentation and outcome presentation, we used a general linear model (GLM) that included conditions at 1) the time of incentive presentation with a parametric modulator of the magnitude of the fixed payment and 2) the time of outcome presentation with a parametric modulator of the deviation of actual earnings from the expected value of reward expectations. Because reward expectations at the time of outcome presentation were a function of both the fixed payment offered and the potential wage rate earnings, we computed the expected value as the probability-weighted sum of the fixed payment and wage rate payment (0.5 × Fixed payment + 0.5 × Wage rate payment).



At the time of incentive presentation, we found that VS activity increased with the value of the fixed payment ($p < 0.005$, cluster-corrected, $k = 187$, peak MNI coordinate = [-16, 10, -6]) (Fig 3A, B), consistent with our hypothesis and previous literature showing that the VS plays a crucial role in signaling increasing prospective reward value (9, 10, 19). Moreover, at the time of outcome presentation, bilateral VS activity increased with the deviation of the actual earnings from the expected value ($p < 0.005$, $k = 20$, peak = [-10, 12, -4] for left, [16, 6, -8] for right) (Fig 3C, D, E). This positive scaling with the deviation between the actual outcome and reward expectations is consistent with previous findings that the VS processes reward prediction error (10–13). Together, our

results shed light on the neural processes underlying expectation-based preferences resulting from effortful activities.

Lastly, we investigated whether neural activity could account for participant-level variability in reference-dependent effort exertion by relating whole-brain neural activity to individual coefficient estimates from a model of effort exertion as a function of fixed payment. We found that individuals more influenced by the fixed payment (i.e., more reference-dependent) showed stronger associations with left VS activity, modulated by the expected value at the time of outcome presentation ($p < 0.005$, cluster-corrected, $k = 143$, peak = [-10, 14, -8]) (Fig 3F, G). Therefore, neural activity in the left VS not only tracked the deviation of actual earnings from the expected value but also explained inter-individual variability in the degree to which external rewards influenced participants' reference-dependent effort exertion.

Discussion

Effortful behavior is often conceptualized as a trade-off between reward and cost (*19–22*), yet such formulations leave open a more fundamental question: *relative to what reference point is effort evaluated?* In economics, reference-dependent preferences demonstrate that labor supply and effort respond not to absolute incentives, but to deviations from expected or target earnings (*1*, *14*, *17*, *23*). However, this work has remained largely theoretical or observational (*2*, *4*, *17*, *18*), leaving unresolved how expectations causally shape effort and how such computations are implemented in the brain. Here, we provide experimental and neural evidence that reward expectations act

as an explicit reference point for effort valuation, shaping both behavior and neural activity during effort-based decisions.

By experimentally manipulating reward expectations independently of outcomes, we show that individuals integrate exogenous expectations when deciding how much effort to exert. Effort provision was systematically biased relative to the imposed reference point, revealing a form of reference-dependent effort valuation analogous to that described in economic models of labor supply (*14*, *18*). Crucially, our design isolates expectations as a causal determinant of effort, rather than an inferred or learned quantity, extending prior economic accounts into a controlled experimental setting. These findings suggest that reference-dependent preferences are not limited to market or labor contexts but reflect a general principle governing motivated behavior.

At the neural level, activity in the ventral striatum (VS) tracked both reward expectations at the time of choice and deviations between expected and realized outcomes. These signals closely parallel reference-dependent value computations observed in economic decision-making (*8*, *12*, *13*), providing a neural substrate through which expectations influence effort. Importantly, VS activity also explained inter-individual variability in the degree of reference dependence, linking neural valuation processes to behavioral differences. Rather than encoding effort or reward in isolation, the VS appears to represent outcomes relative to an internally maintained expectation, consistent with a domain-general role in subjective value computation (*24–26*).

Although participants, on average, exerted more effort than required by the fixed reference point, this pattern differs from that in some prior studies, in which individuals

reduced effort once marginal costs exceeded expected returns (*14*). One likely contributor to this discrepancy is the nature of the effort cost. In our task, effort costs were physical and transient, whereas previous work often imposed irreversible temporal costs (*14*). This distinction suggests that reference-dependent effort valuation may be modulated by the perceived recoverability of effort costs. Future studies directly comparing effort domains may clarify how cost structure shapes expectation-based motivation.

Together, these findings bridge economic theories of reference-dependent preferences with neurobiological mechanisms of motivation. By demonstrating that effort is evaluated relative to expected reward—and by identifying a neural signature of this computation—our results provide a mechanistic account of how expectations shape willingness to exert effort. This framework offers a unifying perspective on effort and valuation, with implications for understanding persistence, fatigue, and motivational deficits across behavioral contexts.

References

1. V. P. Crawford, J. Meng, New York City Cab Drivers' Labor Supply Revisited: Reference-Dependent Preferences with Rational-Expectations Targets for Hours and Income. *Am. Econ. Rev.* **101**, 1912–1932 (2011).

2. B. Bartling, L. Brandes, D. Schunk, Expectations as Reference Points: Field Evidence from Professional Soccer. *Manag. Sci.* **61**, 2646–2661 (2015).

3. D. Gill, V. Prowse, A Structural Analysis of Disappointment Aversion in a Real Effort Competition. *Am. Econ. Rev.* **102**, 469–503 (2012).

4. E. J. Allen, P. M. Dechow, D. G. Pope, G. Wu, "Reference-Dependent Preferences: Evidence from Marathon Runners" (Working Paper 20343, National Bureau of Economic Research, 2014); https://doi.org/10.3386/w20343.

5. C. Gutierrez, T. Obloj, D. H. Frank, Better to have led and lost than never to have led at all? Lost leadership and effort provision in dynamic tournaments. *Strateg. Manag. J.* **42**, 774–801 (2021).

6. D. Kahneman, A. Tversky, Prospect Theory: An Analysis of Decision under Risk. *Econometrica* **47**, 263–291 (1979).

7. T. O'Donoghue, C. Sprenger, "Reference-Dependent Preferences" in *Handbook of Behavioral Economics: Applications and Foundations 1*, B. D. Bernheim, S. DellaVigna, D. Laibson, Eds. (North-Holland, 2018; https://www.sciencedirect.com/science/article/pii/S2352239918300034)vol. 1 of *Handbook of Behavioral Economics - Foundations and Applications 1*, pp. 1–77.

8. B. D. Martino, D. Kumaran, B. Holt, R. J. Dolan, The Neurobiology of Reference-Dependent Value Computation. *J. Neurosci.* **29**, 3833–3842 (2009).

9. J. W. Kable, P. W. Glimcher, The neural correlates of subjective value during intertemporal choice. *Nat. Neurosci.* **10**, 1625–1633 (2007).

10. R. Daniel, S. Pollmann, A universal role of the ventral striatum in reward-based learning: Evidence from human studies. *Neurobiol. Learn. Mem.* **0**, 90–100 (2014).

11. J. Yacubian, J. Gläscher, K. Schroeder, T. Sommer, D. F. Braus, C. Büchel, Dissociable Systems for Gain- and Loss-Related Value Predictions and Errors of Prediction in the Human Brain. *J. Neurosci.* **26**, 9530–9537 (2006).

12. G. Pagnoni, C. F. Zink, P. R. Montague, G. S. Berns, Activity in human ventral striatum locked to errors of reward prediction. *Nat. Neurosci.* **5**, 97–98 (2002).

13. B. Knutson, J. C. Cooper, Functional magnetic resonance imaging of reward prediction. *Curr. Opin. Neurol.* **18**, 411–417 (2005).

14. J. Abeler, A. Falk, L. Goette, D. Huffman, Reference Points and Effort Provision. *Am. Econ. Rev.* **101**, 470–492 (2011).

15. E. Fehr, L. Goette, Do Workers Work More if Wages Are High? Evidence from a Randomized Field Experiment. *Am. Econ. Rev.* **97**, 298–317 (2007).

16. E. Fehr, L. Goette, Do Workers Work More When Wages are High? 500908 [Preprint] (2004). https://doi.org/10.2139/ssrn.500908.

17. H. S. Farber, Reference-Dependent Preferences and Labor Supply: The Case of New York City Taxi Drivers. *Am. Econ. Rev.* **98**, 1069–1082 (2008).

18. B. Kőszegi, M. Rabin, A Model of Reference-Dependent Preferences. *Q. J. Econ.* **121**, 1133–1165 (2006).

19. T. T.-J. Chong, M. Apps, K. Giehl, A. Sillence, L. L. Grima, M. Husain, Neurocomputational mechanisms underlying subjective valuation of effort costs. *PLOS Biol.* **15**, e1002598 (2017).

20. B. De Borger, M. Fosgerau, The trade-off between money and travel time: A test of the theory of reference-dependent preferences. *J. Urban Econ.* **64**, 101–115 (2008).

21. M. Burrell, A. Pastor-Bernier, W. Schultz, Worth the Work? Monkeys Discount Rewards by a Subjective Adapting Effort Cost. *J. Neurosci.* **43**, 6796–6806 (2023).

22. T. Müller, M. C. Klein-Flügge, S. G. Manohar, M. Husain, M. A. J. Apps, Neural and computational mechanisms of momentary fatigue and persistence in effort-based choice. *Nat. Commun.* **12**, 4593 (2021).

23. E. Fehr, L. Goette, Do Workers Work More if Wages Are High? Evidence from a Randomized Field Experiment. *Am. Econ. Rev.* **97**, 298–317 (2007).

24. O. Bartra, J. T. McGuire, J. W. Kable, The valuation system: a coordinate-based meta-analysis of BOLD fMRI experiments examining neural correlates of subjective value. *NeuroImage* **76**, 412–427 (2013).

25. G. Sescousse, Y. Li, J.-C. Dreher, A common currency for the computation of motivational values in the human striatum. *Soc. Cogn. Affect. Neurosci.* **10**, 467–473 (2015).

26. V. S. Chib, A. Rangel, S. Shimojo, J. P. O'Doherty, Evidence for a Common Representation of Decision Values for Dissimilar Goods in Human Ventromedial Prefrontal Cortex. *J. Neurosci.* **29**, 12315–12320 (2009).

Acknowledgments: None

Funding: This work was supported by the Eunice Kennedy Shriver National Institute of Child Health and Human Development of the National Institute of Health under the Award Number R01HD097619 (VSC).

Author contributions:

Conceptualization: AK, VSC

Methodology: AK, VSC

Investigation: AK, VSC

Visualization: AK, VSC

Funding acquisition: VSC

Project administration: AK, VSC

Supervision: VSC

Writing – original draft: AK

Writing – review & editing: AK, VSC

Competing interests: Authors declare that they have no competing interests.

Data and materials availability: Data and codes used in the analysis are available in the project's Open Science Framework website (https://osf.io/t3q72/).

Fig. 1. Experimental design. (A) Experimental schedule. Participants began the experiment by performing a maximum grip exertion, which was measured to normalize each individual's strength for the effort task. They then performed a Training Phase in which they developed associations between the exertion required to obtain different earnings levels. (B) A training trial presented a target earning level, and a participant performed an effortful grip with real-time visual feedback of the exerted force represented as a bar that increased in height with increased exertion. The bar turned green when the participant reached and held the target exertion level. Effort levels ranged from $0 (no effort) to $4 (80% of maximum grip exertion). A Recall Phase (C) was performed to determine whether participants developed salient associations between exertion levels and earnings. During trials, participants were presented with a target incentive level and instructed to exert effort at that level without feedback of exertion. (D) Participants were able to accurately exert to the target earnings levels during the training phase. (E) During the Recall Phase, participants exerted the appropriate effort associated with the presented earning level. (F) During each trial of the Effort Provision Phase, participants were presented with a risky option that would yield either a fixed monetary payment, regardless of the effort they exerted, or a wage payment proportional to the effort they exerted. Each option had an equal probability of occurring, and the fixed payment ranged from $0 to $4 to create different expectations

among participants. The actual outcome, either the wage or fixed payment, was realized after individuals had a chance to exert effort.

Fig 2. Behavioral Results. Participants could take two approaches in deciding how much effort to allocate when presented with the possibility of receiving either earnings proportional to their exertion or a fixed payment. (A) In one case, if an individual felt that the effort was not costly or aversive, they could exert their maximum effort to maximize potential wage earnings and the expected outcome value of a trial, and disregard the level of the fixed payment. (B) In another case, if participants found effort costly and aversive and wanted to avoid exerting effort that might not lead to a wage payment, they could modulate their exertion levels to account for the fixed payment amount, in hopes of reducing their exertion to avoid wasted effort. (C) Effort exertion as a function of the fixed payment during the Effort Provision Phase. Both effort and fixed payment were normalized so that 1 effort unit corresponded to 100 effort units, and 1 unit of fixed payment corresponded to $4. The gray points and error bars represent the mean and standard error of effort exertion across participants for each fixed payment. The black line and shaded area represent the regression fit and confidence intervals across all participants. The dotted line represents the case where participants' wage earnings match the fixed payment. (D) Intercept and slope from the individual regression models. Gray points represent individual participants; the black points and whiskers are the mean and standard error across participants. (E) A histogram of how effort-based wage

earning deviated from fixed payment. A value greater than 0 indicates wage earnings exceeding the effort required for the fixed payment.

Fig. 3. fMRI Results. (A) Whole brain results during the Effort Provision Phase at the time of Incentive Presentation. Activity in the left VS (peak = [-16, 10, -6]; small volume corrected $p < 0.05$ in a priori ROI) was positively modulated by the fixed payment. (B) Effects in the right VS (5mm sphere centered at [-13, 17, -8]) for low ($0.25-$2) and high ($2.25-$4) fixed payments. This was illustrative only and was not used for statistical inference (C) Whole brain results during Outcome Presentation. Activity in the VS (left peak = [-10, 12, -4], right peak = [16, 6, -8], small volume corrected $p < 0.05$ in a priori ROI) reflected the difference between the actual outcome and the expected value for the fixed payment and wage payment, given an individual's exertion. (D) Effects in the bilateral VS (5mm sphere centered at [-6, 7, -13] for left and [21, 7, -10] for right) for low, medium, and high deviations of the outcome from the reward expectations. This was illustrative only and was not used for statistical inference. (E) A histogram of the parametric modulator (Earning – Expected Value) used in the analysis during the Outcome Presentation. (F) Whole-brain covariate results during the Outcome Presentation. There was significant activity in the left VS associated with the degree of an individual's reference-dependent effort exertion (peak = [-10, 14, -8]). (G) Neural activity in the nucleus accumbens was significantly correlated with individual differences in reference-dependent tendency during effortful activity, as estimated from the individual regression model ($r=0.74$, $p<0.001$).