Know When to Fold 'Em: The Flip Side of Grit*

Larbi Alaoui† Christian Fons-Rosen‡

January 13, 2021

Abstract

This paper investigates the way different sides of grit influence behavior. In addition to grit's upside in achieving economic success associated with not giving up, it might also have a downside associated with not letting go. We split grit into two new categories, tenacity and diligence, and hypothesize that tenacity can lead individuals to go beyond their own intended plan of action when making a loss. We test the predictions with an experiment that elicits each individual's plan of action which we compare to actual choice in a game of luck. Consistent with our priors, grittier individuals have a higher tendency to overplay, and tenacity alone captures the difficulty in respecting ex-ante preferences when this means accepting defeat. We then discuss the external validity of our findings.

Keywords: noncognitive skills; grit; tenacity; diligence

JEL codes: C91; D01; I20

---

*Alaoui and Fons-Rosen gratefully acknowledge financial support from the Spanish Ministry of Economy and Competitiveness under project PGC2018-098949-B-I00, from the Spanish Ministry of Science and Innovation under grant RYC-2016-21127 (Alaoui), as well through the Severo Ochoa Program for Centers of Excellence in R&D, grant SEV-2015-0563. We thank the editors, two anonymous referees and Sule Alan, Carlos Alos-Ferrer, Jose Apesteguia, Ghazala Azmat, Roy Baumeister, Roland Benabou, Veronica Benet-Martinez, Laura Doval, Angela Duckworth, Seda Ertac, Rosa Ferrer, Andrea Galeotti, James Heckman, Johannes Horner, Philipp Kircher, David Laibson, Jennifer Lerner, David Levine, Pablo Lopez-Aguilar, Christian Michel, Magne Mogstad, Ignazio Monzon, Antonio Penta, Helena Perrone, Richard Roberts, Santi Sanchez-Pages, Charles Sprenger, Philipp Strack, Adam Szeidl, Ana Tur-Prats, Kathleen Vohs, Leeat Yariv, participants of the HCEO Chicago conference, the UPF behavioral science workshop, the SAET conference, and members of the Bogazici University, the EUI, Sussex university and the University of Pennsylvania Character Lab for their thoughtful comments.

†Universitat Pompeu Fabra and Barcelona GSE. E-mail: larbi.alaoui@upf.edu

‡University of California at Merced, MPI Munich and CEPR. E-mail: cfons-rosen@ucmerced.edu

1 Introduction

The influence of personality traits and noncognitive skills on success has been progressively recognized in recent years (Almlund, Duckworth, Heckman and Kautz (2011)). Among these skills, grit and conscientiousness, which connect to perseverance, have attracted attention in the psychology literature and increasingly in economics. Studies consistently show that such skills are critical in education (see, e.g., Heckman and Rubinstein (2001), Duckworth, Peterson, Matthews and Kelly (2007), Dobbie and Fryer (2015) and Burks et al. (2015)), and may even outperform IQ in determining lifetime success (Duckworth and Seligman (2005), Roberts et al. (2007), Kautz et al. (2014)).

The upside of grit and conscientiousness is clear upon introspection, especially in their link to a notion of perseverance, which is nearly universally accepted as crucial to achieving success. At the same time, the notion of stubbornness may also seem linked, and it has a decidedly more negative connotation. The same characteristics which are so critical for success also seem, when cast under a different light, related to the idea of not knowing when to let go. Little is known about the repercussions of this second aspect, and about whether the effects of the two components, the positive and the potentially negative, can be disentangled using standard measures of grit.[1] Our primary aim in this paper is to investigate whether grit can have a negative component. We find that this may indeed be the case. We also find, consistent with the hypothesis on which we expand below, that one specific dimension in grit, which we denote 'tenacity', appears to be responsible for the negative aspect.

The first task is to define the meaning of not letting go at the right moment, as it is an inherently subjective idea. Within economics, a widely accepted viewpoint for evaluating decision-making consists of using the individual's own ex-ante preferences, or plan of action, as the metric. Under this view, the determining factor is internal consistency, which compares the ex-ante plan of action to the actual behavior exhibited ex-post. But making this comparison using traditional datasets poses a challenge, because the plan of action is typically not observed. Suppose, by way of example, that when playing a game of luck (or investing in the stock market), an individual loses a large amount but refuses to stop. This may be fully consistent with an ex-ante preference of pursuing a risky strategy, but it is also consistent with him deviating from his plan and "overplaying," finding it more difficult than expected to accept a loss. For this reason, we design an experiment that explicitly asks for a plan of action and relates it to the different facets of grit. Noting that the risk-attitude translates to the plan of action, this approach also allows us to separate overplaying from risk attitude.

Our experiment elicits both the ex-ante plan of action and actual behavior using a simple single-agent game of pure chance. This game has no ability component and no scope for overconfidence, and it induces well-defined temptations to go beyond the established plan. In particular, subjects start with a fixed amount of money, which they can win or lose. They can effectively

[1] A number of studies have found a high correlation between grit and conscientiousness (a "Big Five" personality component, together with neuroticism, openness to experience, extraversion and agreeableness), and both are intimately linked to the idea of perseverance. We focus on the grit survey because it includes questions which are more directly phrased in the language of what we term tenacity, but, as we discuss explicitly below, we expect some of our findings to hold with conscientiousness as well.

keep playing for as much or as little as they wish. Their elicited plan of action determines the maximum loss that they are willing to incur before stopping the game, which is determined by their risk attitude. Going beyond the plan, therefore, means potentially incurring a bigger loss. We measure grit using the Duckworth and Quinn (2009) questionnaire. As we will show, grit can be decomposed into two novel components, which we denote as 'tenacity' and 'diligence', where only the former is responsible for overplaying.

Our key hypotheses are that (i) agents with higher grit have a higher likelihood of incurring a cost of failure when they are losing, and (ii) when splitting grit into tenacity and diligence, it is tenacity that drives overplaying, and not diligence.[2] In the context of our experiment, we take losing to specifically mean being below the initial endowment. Under this hypothesis, more tenacious agents have a higher probability of overplaying and going past their planned maximum loss, as they attempt to get out of the loss domain so as to avoid incurring the extra cost of failure.

The results of our experiment are consistent with our hypothesis. Comparing the actual choices of subjects to their plan of action, we find that grittier subjects overplay the most, even after controlling for preferences. At least 30% of subjects overplay, irrespective of characteristics such as undergraduate degree and gender.[3]

Our split of grit into diligence and tenacity is a new decomposition and it serves our objectives because it isolates the central aspect of not letting go (tenacity) from the aspect of working harder (diligence). Consistent with our predictions, only tenacity drives overplaying in the regressions. While diligence plays no role in our experiment, tenacity has a downside.[4] In addition, we perform a two-factor confirmatory factor analysis (CFA) of the main decomposition of grit into diligence and tenacity as well as of the variations used in the robustness checks, and find that the results of this analysis are consistent with our categorization. This result also supports the method employed by Alan, Boneva and Ertac (2019) who, introducing a behavioral measure of grit, find that it is multidimensional.[5]

We do note, however, that we do not directly measure risk aversion, other than through the plan of action. While the plan of action aims to capture the agent's risk preferences, and agents with higher tenacity are not typically associated with being more risk averse, the lack of direct measure of risk aversion suggests some degree of caution in interpreting our results. Moreover, there may be characteristics that we we do not observe that correlate with grit and which may explain our findings. For instance, suppose that subjects with higher tenacity are also worse at assessing how enjoyable the roulette game will be. This could then lead to overplaying, and explain our findings. While we favor the interpretation provided in this paper, we stress that

[2]Since our experiment is designed to remove considerations such as hard-work, diligence should play no role, as we will test.

[3]This illustrates that our design can be useful for future experiments which require a domain in which temptation is easily elicited and measured. As discussed later on, this domain of temptation is relatively unexplored in the literature.

[4]In other settings, one can imagine that diligence also has a downside and an upside of a different form- the upside being that hard work is useful for productivity, but overworking may be damaging to the agent. We thank a referee for this observation.

[5]The behavioral measure introduced by Alan et al. (2019) involves eliciting various aspects of grit using a new incentivized task.

future research is needed to rule out alternative explanations such as the one proposed here.

The tenacity and diligence categories are available in any dataset that uses the standard grit questions, and so they can easily be used in future research. As a simple illustration, we analyze existing data on educational performance. Education is a domain in which we would conjecture that the positive aspect of tenacity dominates. Online Appendix C shows that the tenacity and diligence measures are both positive predictors of educational level. However, we stress that this analysis is mainly exploratory and indicative at most, since, unlike the data obtained from our experiment, we do not have the ex-ante plan of action.

Looking ahead, the finding that grittier individuals may have a higher tendency not to let go when they should, even according to their own preferences, raises policy implications and questions for future research. Our results indicate that it is specifically tenacity that should be investigated for the potential downside identified here. For instance, it is well-known that investors often have difficulty in accepting their losses (e.g. Odean (1998)), and numerous firm managers and entrepreneurs maintain their strategy despite all indications that they should readjust. Future research can focus on the interplay between the different facets of grit and conscientiousness as functions of the environment. As mentioned above, however, future research should also account for the interaction between tenacity and risk aversion.

The rest of the paper is structured as follows. We first discuss the related literature. Section 2 then presents the experiment and the predictions, Section 3 provides the main findings, and Section 4 discusses additional results. Section 5 addresses the external validity of decomposing grit into tenacity and diligence. Section 6 concludes.

Related literature

This paper relates to several strands of research. The first concerns the relevance of noncognitive skills, and particularly grit and conscientiousness, to economic outcomes such as education and employment. The psychology literature is more extensive while the economics literature is small but growing. Here we do not provide an exhaustive survey, but see Borghans, Duckworth, Heckman and Weel (2008) and Almlund et al. (2011) for detailed discussions. Grit is especially important for achieving long-term goals, even in the absence of positive feedback (Duckworth et al. (2007)). Heckman and Rubinstein (2001) explain the importance of noncognitive skills in academic success, and highlight the gap in the economics literature in analyzing them. MacCann, Duckworth and Roberts (2009) find that the different facets of conscientiousness are linked to higher education performance, and Burks et al. (2015) find that within conscientiousness, the 'hard work and persistence' component predicts collegiate success. Dobbie and Fryer (2015) consider the impact of charter schools on outcomes such as academic achievement, and among their measures, administer the short-scale Duckworth Grit Index used in this paper.

Gerhards and Gravert (2017) propose a behavioral measure of grit. Cubel, Nuevo-Chiquero, Sanchez-Pages and Vidal-Fernandez (2016) conduct a laboratory experiment on the effect of the Big Five on performance, and find that conscientious subjects perform better. Gill and Prowse (2016) analyze how both cognitive skills and personality, including grit, influence behavior in the beauty contest. Proto, Rustichini and Sofianos (2018) study how intelligence and personality affect group outcomes, and find that conscientiousness and agreeableness have an impact on cooperation. Papageorge, Ronda and Zheng (2017) study how another noncognitive skill, known as externalizing behavior, reduces attainment but is productive in the labor market. Within neureconomics, Rueter et al. (2018) test the hypothesis that a neural network denoted as the "goal priority network" is a neural correlate of conscientiousness. Within the psychology literature, Lucas et al. (2015) find that grittier subjects are less willing to quit when failing. The focus of their study is distinct from ours, however, and since there is no measure of the plan of action, the subjects' intentions cannot be compared to their actual behavior. Boyce, Wood and Brown (2010) use data from a German socio-economic survey and find that the negative correlation between unemployment spells and life satisfaction is stronger for individuals with high levels of conscientiousness. Using another survey, Heineck (2011) finds a non-linear association between conscientiousness and wages.

To some extent our analysis using grit is likely to have implications on conscientiousness as well, since recent studies find that they are intricately related. In particular, Credé, Tynan and Harms (2016) provide a meta-analysis that finds a high correlation between grit and conscientiousness, particularly with respect to the 'perseverance' facet of grit. The focus of our analysis is orthogonal to establishing the relation between grit and conscientiousness, but an interesting subject for future research concerns the extent to which our results carry through using the conscientiousness measure. We discuss this point in Sections C.2.

Considering the importance of noncognitive skills for success, a natural policy issue concerns the degree to which these skills can be influenced. Cunha, Heckman and Schennach (2010) analyze the production function of both cognitive and noncognitive skills, in a dynamic setting in which skills are determined through parental investment, and estimate optimal targeting of interventions. Recent papers use large-scale policy interventions in elementary schools to investigate the malleability of noncognitive skills. Alan and Ertac (2014) find that groups treated towards being more patient have improved outcomes, even a year after the intervention. Their findings strongly suggest that noncognitive skills can be influenced in a persistent manner. Alan, Boneva and Ertac (2019) focus more explicitly on a notion of grit, and also find that treated subjects are more likely to obtain higher grades, using a novel incentivized real effort task.

We view this paper as complementary to the ones above. By revealing another side of grit, that of not letting go, our paper shows that grit is not only a prerequisite for achievement, it can also shed light on other patterns of behavior. In particular, it can help explain the difficulty in respecting one's own plan of action and accepting a loss. Our paper also shows that the relevance of grit can be elicited in a laboratory environment and in a short timespan.

A second strand of the literature concerns dynamic inconsistency in decision making, which occupies a central part in behavioral economics. The most commonly used modeling approach is to use present-biased preferences. The reach of this approach is too vast to document in this paper, but see, among other contributions, Strotz (1956), Laibson (1997), O'Donoghue and Rabin (1999); see also O'Donoghue and Rabin (2015) and Sprenger (2015) for discussion of the literature; for other modeling approaches, see, e.g., Gul and Pesendorfer (2001), and see Ameriks, Caplin, Leahy and Tyler (2007) for a survey-based measure of self-control problems. An important question within this literature concerns separating ex-ante preferences from behavior and measuring the extent of present-bias. Augenblick, Niederle and Sprenger (2015) resolve the main difficulties of this exercise in a laboratory experiment, over a span of several weeks. For papers that focus on commitment, see, for instance, Ashraf, Karlan and Yin (2006) on commitment devices for savings products, Kaur, Kremer and Mullainathan (2015) for a field experiment on self-control and commitment devices in the workplace, and Alan and Ertac (2015) for a field experiment that analyzes patience, self-control and the demand for commitment among school children. Recently, an interesting paper by Heimer et al. (2020) analyzes dynamic inconsistency in risky settings in the lab and the field. It finds that subjects deviate from their strategy by chasing losses and cutting their gains early.

Our paper does not focus on present-bias preferences, but the connection is in the dynamic consistency problem itself. Grit is sometimes linked to – though distinct from – self-control ability and self-discipline (see Duckworth and Gross (2014)). We also note that Amador, Werning and Angeletos (2006) consider the optimal tradeoff between commitment and flexibility. Our paper has a different focus and we do not test whether grittier subjects are more or less likely to commit to a goal beforehand or to prefer flexibility. Rather, we analyze whether they are more likely to systematically go beyond their pre-established plan.

Our setting also has a self-control element, but the temptation is very different in nature. In this sense, our paper brings to light a relatively unexplored domain of temptation namely that of not letting go and accepting failure (but see, within the psychology literature, Gollwitzer (1999) and Achtziger, Bayer and Gollwitzer (2012)). This form of temptation may relate to noncognitive skills in a way that is distinct from conventional domains. Grittier subjects report less self-control problems, in line with known results, but within our game they are also more likely to deviate from their own plan of action. Another contribution of our paper is methodological. At least 30% of subjects overplay, illustrating that our design can be useful for experiments that require a domain in which temptation is easily elicited in a short period of time.

A separate and extensive literature analyzes the effects of overconfidence (see, for instance, Malmendier and Tate (2005), among many contributions). Our experiment has been designed to avoid overconfidence effects, in that there is no ability component on the subject's part, and the probabilities of failure are known and objective. However, an interesting direction for future research may involve exploring the relation between the grit and tenacity of CEOs and their overconfidence. For instance, Galasso and Simcoe (2011) find that more overconfident CEOs, i.e., those who underestimate the probability of failure, are more likely to pursue innovation. Also aligned with this view is the paper by Barbosa, Gerhardt and Kickul (2007) who have a sample of 528 entrepreneurial students to examine how both cognitive style and risk preference contribute to one's assessment of abilities and entrepreneurial intentions. They find that individuals with a high risk preference have higher levels of entrepreneurial intentions whereas individuals with a low risk preference rather had higher levels of relationship efficacy and tolerance efficacy. In another interesting study, Sapienza, Zingales, and Maestripieri (2009) have a sample of over 500 MBA students to investigate whether between- and within-gender variation in financial risk aversion was accounted for by variation in salivary concentrations of testosterone. The authors find that testosterone has a nonlinear effect on risk aversion regardless of gender. Interestingly, both testosterone levels and risk aversion predict career choices after graduation: Individuals high in testosterone and low in risk aversion were more likely to choose risky careers in finance.

Within the prospect theory literature, Barberis (2012) discusses how it can explain gambling behavior, and particularly the idea of gambling past the plan to follow a particular strategy. Ebert and Strack (2015) show that, strikingly, a naive agent with prospect theory preferences will gamble until the bitter end, while Ebert and Strack (2018) show that a sophisticated one will not start gambling at all.

It would be interesting to link tenacity to higher loss aversion through a channel that relates to prospect theory. The focus of these papers is distinct from ours, which focuses on comparative statics on overplaying along the grit (and tenacity) dimension. While using a prospect theory setting would be interesting, we do not explore whether tenacity links directly to loss aversion (in a prospect theory sense) or probability weighting. This is partly because it would be challenging to explain through such a channel why, in our experiment, tenacious subjects do not have significantly different plans of action and do not play more than others; rather, they *overplay* more. In addition, Ebert and Strack's (2015, 2018) results suggest some caution in this regard.

Lastly, there is a large literature on the disposition effect, which refers to investors' tendency to keep assets whose value have dropped (see, e.g., Shefrin and Statman (1985) and Odean (1998)). This effect may be present in the way subjects approach losses in our experiment. For papers exploring the disposition effect combined with a plan of action, but without incorporating grit or tenacity, see Andrade and Iyer (2009), Ploner (2017). Relatedly, Fischbacher, Hoffmann and Schudy (2017) investigate how to mitigate the disposition effect via automatic trading systems. In light of our results, another avenue of research would explore whether more tenacious subjects also have a higher tendency to hold losing assets longer than they had planned on ex-ante.

2 Experimental Design

This experiment was conducted at the Behavioral Sciences Laboratory (BESLab) of Pompeu Fabra University, and included 138 subjects. There were eight sessions spread over several weeks, and typically lasted between 75 and 90 minutes, with 15 to 20 subjects per session. Each session is divided into three main stages: the first consists of eliciting the subjects' plan of action for the (single-agent) games in mind, the second consists of subjects actually playing the games, and the third consists of a questionnaire. We explain each stage below, after describing the main game that subjects can play.

2.1 Description of the game

The relevant game for our analysis is a simplification of the familiar (American) roulette. We modified a standard online roulette by removing some betting options (e.g., betting on evens and odds and on first, second or third 12th). For screenshots of the roulette, see Figure 1, and for the actual game used see http://experimentalgames.upf.edu/roulette/.

Each subject has 2000 tokens to start with, where 150 tokens is worth 1 euro. He or she (henceforth he) can place a bet on any of the 38 numbers (0, 00, and 1 to 36) or colors (red or black), or on nothing at all. He can bet anywhere between 0 and 500 tokens per spin, in any increment of 50. Once he finishes placing his bets, the roulette wheel spins, and the ball lands on a number and its associated color. We use the standard roulette betting rules: if the ball lands on a number on which the subject has placed a bet, he receives the original amount wagered on the number plus 35 tokens for each token wagered. If it lands on a color on which he has bet, then he receives the original amount plus 1 token for each token wagered. All other tokens bet are lost, and the game repeats with the updated amount of tokens. Subjects can play for as long or as little as they wish, and they do not have to play if they do not want to. They can also spin the wheel even when their bet is 0. Subjects can see the tokens they have, the amount bet, and the history of the last five roulette outcomes (color and number). They cannot earn less than 0 tokens or more than 8000. Notice that the expected returns are negative for any strategy that involves betting a positive amount.

We use this game because it has several advantages for our objectives. First, this game is familiar to many, so subjects do not incur a large cognitive cost to understand it. This notion was confirmed during the experiment, in which almost no subject displayed problems understanding our roulette. Second, there is no ability component. This is a game of pure luck, and furthermore each draw is independent of the others. This feature reduces any potential confounds from overconfidence in one's own ability, which may otherwise form before or during the game.[6] Third, this game quickly draws the subjects in, absorbs their attention, and makes them value the outcome. While it has a component of entertainment, it simulates an environment in which losing and winning are important to the subjects. The amount they receive is itself meaningful, in that their potential earnings (8000 tokens, or 53 euros) are considered substantial in Spain, as is the magnitude of the loss from their initial wealth to 0. For example, with 53 euros one could pay around 10 canteen meals at university or 8% of the average monthly rent in the city center. Alternatively, if one divides the average gross monthly salary by 20 days worked in a month, one obtains that the average daily wage that is almost the amount of money they can win in our experiment. This can easily be observed during the experiment from the subjects' reactions to their outcomes. Lastly, even if subjects play less or more in an experimental setting than they would otherwise, this would not introduce any bias in favor of our findings. That is, it should not lead grittier individuals to have a higher tendency to deviate from their plan of action.

2.2 Plan of action

The first stage of the game consists of eliciting the subjects' plan of action.[7] We inform them of the game rules and of their initial endowment of 2000 tokens. We then ask each subject for the range of tokens within which he would like to play. Specifically, we ask for the minimum

[6] The data shows that subjects do not use gambler's fallacy strategies, so this is not an issue. Moreover, such behavioral biases would not correlate with grit or other critical factors for our analysis.

[7] The experiment was programmed in Qualtrics, aside from the roulette itself which was programmed in Adobe Flash.

limit which he is not willing to surpass. (For completeness we also ask for the maximum of the range, but it is not the object of our analysis.) This limit is the subject's intended bound, in that his plan is to stop playing if it is ever reached or even before. He can choose any number in increments of 100 between 2000 tokens and 0 tokens. That is, he can choose within the range of not playing at all and risking his entire wealth. We refer to the elicited lower bound as the "planned minimum bound," or simply planned minimum.

We believe the subjects to be truthful about their plan of action. They have no incentives to be dishonest, and lying is well-known to be psychologically costly (see, for instance, Gneezy (2005)). Moreover, any noise in the responses should not correlate with grit, and would therefore not bias our results.

As an additional precaution, however, our experiment consisted of a first game that subjects play before the roulette. Using this game serves to check for the reliability of the stated plan of action. Specifically, before introducing the roulette game and asking for the plan of action, we ask the subjects for their plan in this first game. We do not expect to observe a high degree of overplaying in this first game, as it is designed to be much less captivating and tempting. Comparing the stated plan of action to actual play in the first game then provides a measure of how accurate elicited plans are. In particular, finding that the stated plan of action and the actual play are close in the first game would support the view that subjects' stated plan of action is reliable for the roulette game as well, since there is no reason for subjects to be less truthful in the roulette than in this first game.

The rules of the first game are the following. Subjects are informed that they will start with 2000 tokens, just as they will later be informed of the same initial endowment for the roulette. At each round of the game, he chooses a number between 1 and 36. The computer chooses a number at random in that range as well, and if they match then he receives 3200 tokens and the game stops automatically. Otherwise, he loses 100 tokens. Here too, the subject can play for as many rounds or as few as he wishes, unless he reaches 0 tokens or unless he wins. We ask for the subject's plan of action, and specifically for the number of rounds that he would like to play, provided he has not won before. Since the game ends if the subject wins, his maximum earnings are therefore 5200 tokens, which he would receive if he wins in the first round. This game was designed to be much more passive than the roulette, since the only choices available here are whether or not to stop at any stage and to pick one of the 36 numbers. Furthermore, since the game ends as soon as the subject wins once, there can be no post-victory momentum from playing after a win takes place.

As discussed above, we do not expect the subjects to lie about their plan of action. But to reinforce their incentives to tell the truth, we use a variation of the Becker-DeGroot-Marschak (BDM) method to make this stated plan of action incentive-compatible. We leave the details of this point to the Appendix, as they are involved and not central to our analysis. In brief, however, at the start of the experiment we explain to the subjects the BDM mechanism and its incentive compatibility. After asking for their stated plan of action, we ask for their willingness to sell their preferred plan of action in exchange for having to play until they either lose all their tokens or win the maximum amount possible. We explain that we may compare their chosen amount against an amount chosen by the computer, according to the BDM (second-price auction) way.[8] We are not using this amount in our analysis, but only the stated plan of action, which is the focus of our study. The results from the first game suggest that the subjects are indeed truthful in their response and wish to respect their established plan of actions. The reason we do not use the BDM is that, while we think that it does incentivize the subjects to give the plan of action and that it is not too cognitively challenging (which is consistent with the vast majority of subjects answering the BDM comprehension questions correctly), estimating the true willingness to pay (or sell) this plan is a much more complex task, and likely to be noisy. That said, in Section 4.3 we do mention an additional result that uses the BDM, although it is not part of our main analysis.

Each subject's final payment is based on the earnings of either the first game or the roulette, chosen at random. We inform the subjects beforehand, explaining that at the end of the experiment a die will be tossed to determine the earnings. We do so to avoid any wealth effects, as may occur if the final earnings were based on the sum of the earnings in the two games. This ensures that the games are independent and that earnings in one have no financial impact on earnings in the other.

After eliciting the subjects' plan of action in the first game and in the main (roulette) game, we ask additional questions with the aim of attenuating any anchoring effect from the recently elicited plans of action before they play the actual games. These questions, which take approximately 10 minutes to answer, are designed not to impose a heavy cognitive burden or induce fatigue. They consist, for instance, of filling in a caption for a drawing and of checking boxes on the kinds of news that they follow. The subjects are informed that these questions will not impact their earnings in any way. They are free to answer them as they wish, or even randomly if they are inclined to do so, without any consequences.

2.3 Game play

After subjects provide their plan of action and answer the questions that follow, they play the two games. There are no time constraints of any kind, and each subject's game is private and independent of everyone else's. Each subject plays the games in the same order as they were presented in the 'plan of action' stage: he starts with the first game presented (the simpler 1 to 38 game), and then plays the main one (the roulette).

2.4 Questionnaire

The subjects are asked to complete a 5 minute questionnaire in the last stage for an additional 600 tokens. This survey includes the standard 8-question short grit measure (see Duckworth and Quinn (2009)). These questions are used to construct the Grit Index, which ranges from 1 to 5,

[8] We specify in the description that we 'may' use the BDM because this choice was not made binding, in that all subjects were allowed to play the lottery. The instructions were specific on this point, and no subject expressed confusion on this at any stage of the experiment. We also note that we ask three comprehension questions of the BDM, and find that 131 of 138 subjects answer at least two questions correctly.

where grit is increasing in the number assigned.[9] We also ask questions from a shortened Locus of Control (Rotter (1954)), and construct a Locus Index that will be used as a control variable. The Locus Index ranges between 0 and 1, where a higher score is associated with a higher belief that external factors determine events and outcomes. Subjects are then asked additional questions, particularly on their self-assessed degree of temptation and procrastination problems, as well as self-esteem. We specifically use the single-item measure of self-esteem that consists of the question "I have high self-esteem" which ranges from 1 to 5, where 5 indicates higher self-esteem. This measure is viewed as being highly correlated with the Rosenberg Self-Esteem Scale in adult samples (Robins, Hendin and Trzesniewski (2001)). Lastly, subjects were asked to indicate their age, gender and field of study.

2.5 Main hypotheses

Our main hypotheses are presented in this discussion. A formal model is available upon request. Suppose that an agent is endowed with tenacity $T \geq 0$ and diligence $D \geq 0$, which we assume are not negatively correlated. Grit G is the weighted sum of the two (i.e., $G = \alpha T + (1-\alpha)D$, for some $\alpha \in (0,1)$. By way of illustration, consider the image of a person with a rowboat that he can both row and steer. Together, these two factors determine grit, and separately they determine diligence and tenacity. A more diligent person rows harder, and a more tenacious one is more reluctant to steer away from his trajectory, regardless of the setbacks encountered along the way.

Let $z_b > 0$ be the agent's initial endowment, which is 2000 tokens in our experiment. Let $z_m \in [0, z_b]$ be the planned minimum bound, which is the threshold which ex-ante he is unwilling to play lower at when losing. If the agent crosses this threshold –that is, if goes below this amount when playing– then we say that the agent is *overplaying*.

Our central assumptions are that (i) when an agent playing the roulette is losing, meaning when his current wealth is below z_b, he incurs a high unanticipated cost of failure $c_f > 0$ with some probability $q_f(T) \in [0,1)$, that (ii) this probability is increasing in tenacity T ($q_f(\tilde{T} > q_f(T)$ if $\tilde{T} > T$), and that (iii) the Grit Index (from the grit questionnaire in the experiment), and it's decomposition into the Tenacity and Diligence indices, are indicative of a subject's Grit, Tenacity and Diligence, respectively. Under these assumptions, we make the the following three predictions.

Prediction 1: The probability of overplaying (weakly) increases in the Tenacity Index, and does not change in the Diligence Index, except through its possible correlation with the Tenacity Index.

This first prediction will be tested in Section 3.3.

Prediction 2: The probability of overplaying also increases in the Grit Index.

[9] As a robustness check, we run three further sessions in which we ask the subjects these questions at the start of the experiment. As we will discuss later, we do not find differences in the distribution in the answers to the grit questions or its decomposition (using a Mann-Whitney test). Furthermore, the percentage of overplaying in the roulette game differs by less than 1%, and our main results hold with these additional sessions as well. We discuss below the sessions in which the questionnaire is administered at the end, in case placing questions at the beginning makes them salient and affects behavior, but there does not seem to be an issue.

This prediction follows from Prediction 1 and from attribute G consisting of both T and D, using the assumption that the two are not negatively correlated. It will be tested in Section 3.2. Notice that for both predictions, the comparisons are made between two individuals who have the same plan of action but different levels of tenacity. Since different individuals may have different utility functions u, it is important to control for the plan of action, and specifically the planned minimum bound, in the analysis. In particular, our empirical analysis (Section 3) will estimate how the likelihood of overplaying varies with the Tenacity, Diligence and Grit Indices, controlling for the minimum bound. Our prediction, driven by the assumption $q'_f(T) > 0$, is that the sign of the coefficient is positive for the Tenacity and Grit Indices. The sign of the coefficient for the Diligence Index should be 0 once we control for the Tenacity Index.[10]

Prediction 3: Among agents who overplay, the ones with a higher Tenacity Index are more likely to end with final earnings of either 0 or an amount greater or equal to initial earnings z_b.

Here as well, the same link would hold for the Grit Index, accounting for the planned minimum bound. This prediction is more difficult to test, however, due to the highly stochastic nature of final earnings. In closing this section, we note that our predictions can be viewed as providing a mechanism for chasing losses, and specifically that all else being equal, subjects with higher tenacity will be more likely to chase losses (for an interesting recent paper that analyzes chasing losses behavior, see Heimer et al. (2020)).

3 Results

3.1 Descriptive statistics

Since the roulette is a game of chance, a lucky subject may never be close to his minimum limit, and so may never be confronted with the temptation to overplay. Hence, we only capture a lower bound on overplaying. Nevertheless, we still find that 48 of our total of 138 subjects overplay (35%). The high figure illustrates that our roulette is an effective tool for eliciting temptation preferences in a short span of time, in a laboratory setting, and in an easily measured way. Moreover, overplaying is neither gender-specific nor degree-specific; 31% of women and 40% of men overplay, and 34% of those pursuing technical degrees and 36% of those pursuing non-technical degrees overplay. Of the 48 subjects who overplay, 20 of them lose their entire endowment, ending up with 0 tokens. All 138 individuals in our sample chose to play. When measured in terms of the total number of spins played in the roulette, the minimum number of spins is 1, the 25th percentile is 8, the median is 16, and the mean is 28.04, with a maximum of 197 spins.[11]

Table 1 provides further descriptive statistics. The mean for the Grit Index is 3.38, which is in line with other studies. By way of comparison, the mean in Duckworth and Quinn (2009) is 3.4. An average subject plans to stop playing when left with about 900 tokens. In addition,

[10] In the following sections we use the terms tenacity, diligence and grit interchangeably with the Tenacity, Diligence and Grit Indices, respectively, when it is clear that it is a reference to the empirical measures.

[11] The proportion of overplaying in the second game (35%) is statistically larger than the one of overplaying in the first game (16%).

60% are female and the average age is close to 22.

Table 2 takes the same variables already present in Table 1 and measures their correlation with the Grit Index and its decomposition into the Tenacity and the Diligence Indices. Overplaying, our main dependent variable of interest, correlates positively with the Grit Index (0.29) and especially so with the Tenacity Index (0.33). Also consistent with prior beliefs, the Locus of Control index correlates negatively with the Grit Index (-0.19). Finally, self-reported measures of Procrastination and Temptation correlate strongly and negatively both with the Grit Index and with its facets of tenacity and diligence.

Figure 2 presents the graphical distribution of the number of rounds that each subject had planned to play. The Plan of Action ranges from 0 to 2000 with a major spike at 1000, and minor spikes at 500, 1500, and 2000. Figure 3 displays the histograms of the number of rounds each subject overplayed, i.e. went beyond the Plan of Action. Panel A presents it for the full sample of 138 subjects while Panel B limits the sample to the 48 subjects who actually overplayed.

3.2 Overplaying and grit

Overplaying is defined as a dummy whose value is 1 if the subject goes below his planned minimum bound, and it is 0 otherwise.[12] Table 3 presents the first test of our predictions. We begin with the hypothesis that overplaying increases with the aggregate Grit Index (Prediction 2). In column (1), we use a standard OLS econometric specification to regress the dummy for overplaying on the Grit Index. The estimated coefficient is positive, statistically significant at the 1% level, and economically relevant: a one unit increase in the index is associated with a 25.9% increase in the mean predicted probability of overplaying. As individuals with different grit levels also tend to have differing elicited preferences on when to stop playing, our estimated coefficient might be biased. A preference for playing less, which is defined as having a higher planned minimum bound, may lead to a higher likelihood of overplaying, since this bound may be reached with fewer spins. For this reason, we control for preferences (plan of action) in column (2), and find that its estimated coefficient is positive, as expected. Furthermore, the estimated coefficient on the Grit Index is practically unchanged. In columns (3)-(5) we gradually augment the specification with variables controlling for the log of age, a dummy for gender, and a dummy for whether they studied a technical undergraduate degree.[13] The estimated coefficient on the Grit Index remains stable and strongly statistically significant.[14]

In the Appendix we show that results are robust to a battery of tests. In Appendix Table 14 we gradually drop subjects with the largest gains to ensure that they do not drive the results.

[12]Recall that we use the terms 'plan of action' and 'planned minimum bound' interchangeably to refer to the lower limit which the subject does not plan on exceeding.

[13]We define technical degrees as: Economics, Business Administration, Engineering, and Biology. The non-technical degrees are: Political Science, Law, Criminology, Humanities, Sociology, Marketing, Language Translation, Audiovisual Arts, Design, Tourism, Journalism, and Sociology.

[14] In the Appendix Table 13 we perform a robustness check by using an expanded measure of overplaying, which we refer to as 'potential overplaying'. In addition to the 48 individuals to actually overplay, we also include the 8 additional ones who placed a bet that would have brought them below their desired minimum, had they lost. For instance, if an individual's desired minimum is 1000 tokens and he places a 200 token bet when he has 1100 tokens remaining, then he has the potential to go to 900, which is below his desired minimum. We therefore would include him in the potential overplaying measure whether or not he subsequently loses. Results are practically unaffected for this expanded measure.

We then verify in Appendix Table 15 that results are not driven by subjects with little interest in becoming involved in the game. Such players have a high planned minimum bound. For instance, if the planned minimum is 1900 tokens, then the subject does not wish to go beyond 100 tokens below his original wealth of 2000 tokens. The estimated coefficient on the Grit Index remains strongly statistically significant with values around 0.25 in all specifications.

To give visual intuition for our findings, we graphically compare the kernel density estimation of grit for subjects who overplayed and those who did not (Figure 4). This comparison reveals a striking difference between the two, in that the distribution of those who overplayed is markedly shifted to the right. In fact, the cumulative distribution of those who overplayed effectively first order stochastically dominates the distribution of those who have not (Figure 5).

In summary, the finding that subjects with higher grit are more likely to overplay is highly statistically and economically significant, and it is not driven by outliers. Note that this result would have been missed had we not separated ex-ante preferences from actual behavior. Our experimental design, therefore, serves to emphasize the importance of separating plan of action from behavior to address our questions of interest.

Difference in present bias?

Overplaying is a form of dynamic inconsistency in that there is a difference between ex-ante plan of action and actual behavior. The question then arises as to whether these results suggest that grittier agents are more impatient. But since our experiment takes place over a short span of time, temporal preferences should not play a role under any feasible parameters. Moreover, results on temporal preferences and noncognitive skills do not suggest that grittier subjects would be more present-biased (see Alan and Ertac (2015) for a discussion of noncognitive skills and patience).[15] We also do not find any difference in amount of time played, as would have been the case if receiving their final earnings earlier affected subjects in a way that varied with grit. This is in line with the intuition that the form of dynamic inconsistency captured in our environment, which links to difficulty in accepting defeat, is conceptually different from that due to present-biased preferences– in fact, it is possible that they go in opposite directions.

As a robustness check, however, we use the self-reported measures on self-control problems (procrastination and temptation) as controls in Appendix Table 16. Results remain unchanged as the Grit Index is still positive and statistically significant with a similar magnitude.

Possible confounds: risk preferences

We do not directly measure risk attitude in this experiment, and so one may wonder whether differential risk attitude by tenacity could be a driver of our findings. In principle, the subjects' risk-aversion in this experiment is captured through their plan of action– how low or high they set it is determined by their risk attitude. An agent who is fully risk-averse in this setting (in the classical sense of having a preference for the expected value of a lottery to the lottery itself) would set the plan of action to have a minimum bound of 20, meaning that he would not play at all,

[15] This is further confirmed by the subjects' behavior in the first game (the simpler 1 to 38 game) in which grit does not predict any detectable difference in behavior, which would have been the case if deep temporal parameters had been significantly different.

and a more risk-loving agent would set a lower planned minimum bound.[16] Since overplaying is measured with respect to this plan of action, this suggests that the results would not be affected by a direct measure of risk aversion. Moreover, higher grit is not typically associated with less risk-aversion in the literature. As an added check when testing our predictions, in Section 3.3 (Table 7) we split our sample by gender, which is often viewed in the literature as differing in risk attitude. Our results hold for this split as well. But while these arguments are suggestive, having a direct measure of risk attitude in future experiments would be useful in providing added confirmation to the interpretation provided here of the findings.

Additional robustness on main specification

Finally, we produce four robustness tests that confirm that our main finding is not driven by a small subset of observations or outliers. First, we drop subjects who overplayed by a small amount. If our results were solely driven by them, then our claim that grittier people have a higher tendency to overplay would be weakened. In Appendix Table 17 we drop the 10 subjects who overplayed by 100 tokens or less. As before, our main coefficient of interest is unaffected.

Second, in Table 4 we split the sample by technical and non-technical degree in columns (1) and (2) and by gender in columns (3) and (4), while maintaining the remaining usual controls. The estimated coefficients range from 0.201 to 0.312 and the largest values can be found for non-technical subjects and female individuals. While we state that our experiment is purely a luck game in which ability should not play a role, one might wonder that judging optimal stopping time will require the ability to judge the odds and calculate expected returns. One can make the plausible assumption that students choosing technical degrees are likely to have a higher analytical ability. Indeed, we find a slightly lower estimated coefficient for them compared to the subset of individuals enrolled in non-technical degrees.

Third, we consider alternative econometric specifications in Appendix Table 18: Logit, Probit, and Poisson estimators. The estimated coefficient on the Grit Index is again statistically significant at the 1% level throughout the six columns.[17]

Fourth, in Appendix Table 20 we add the Locus of Control Index as a control variable in the OLS regressions and the coefficient has the same magnitude and level of significance.

Lastly, recall that we have asked BDM comprehension questions, which 131 of 138 subjects answered at least 2 of the 3 questions correctly. We re-estimate Tables 3 and 6 by removing the 7 subjects who answered at most 1 question correctly, and find that our results are comparable to our baseline results (these tables are available upon request).

3.3 Decomposing grit into tenacity and diligence

Our explanation for the main finding is that grittier subjects, by being more tenacious, find it difficult to let go and accept failure, even if it means deviating from their initial plan of action.

[16] Of course, the reason for a difference in risk-lovingness (in the standard sense described above) may be due to enjoyment of the game in this context, which is why we would not take this experiment as indicative of risk-aversion in other settings.

[17] In the Appendix Table 19 we perform an additional robustness test where we rerun the regressions of Appendix Table 18 with 'potential overplaying', the expanded measure of overplaying defined previously (Footnote 14), and results are again practically unaffected.

To test this hypothesis, we partition the grit questions into two new categories, 'diligence' and 'tenacity'. These two components are correlated but distinct.[18] They map to the variables D and T, respectively, in Section 2.5. According to Prediction 1, tenacity alone should explain overplaying. Intuitively, recalling the image of the rowboat, the unwillingness to accept defeat and stop playing once the planned minimum bound is reached is a refusal to steer away. It therefore falls squarely within tenacity, not diligence.

Since this decomposition of grit into tenacity and diligence is novel, a concern may be whether it has been done ex-post. We explain here our rationale behind this decomposition and our robustness checks. The mechanism behind our hypotheses operates through the difficulty in letting go, in the sense that some individuals have a higher likelihood of incurring a cost of failure (which is what we denote tenacity). Our main focus, therefore, was to identify questions within grit that were specifically about letting go, and we classified those as tenacity. Since the remaining questions were about being hard-working, we classified those as being about diligence. For instance, the question "Setbacks don't discourage me" clearly falls within tenacity, while "I am diligent" obviously falls within diligence (see Table 5). As with the Grit Index, the Tenacity Index and the Diligence Index can take values between 1 and 5, where 5 indicates higher tenacity and higher diligence, respectively.

To alleviate concerns on this decomposition, we conduct further analysis below, which we now briefly summarize: First, we conduct a confirmatory factor analysis (CFA) to confirm that this categorization is robust. Second, we provide an alternative categorization, where we only use questions that appear to be as clearly about tenacity and diligence as we could reasonably expect, and discard all the others. Third, we show that our categorization, which is closer in spirit to our hypotheses than the standard decomposition of girt, also performs better in this setting. Fourth, we analyze two additional, external datasets in Section 4.6 to test whether tenacity and diligence have any explanatory power on the outcome variables, and find that they do. Taken together, our categorization appears to be the appropriate one for the mechanism we describe here. We note that any dataset which contains the grit questionnaire could be decomposed into these categories, and so future research can further test the extent to which this decomposition is useful.

Table 6 presents the results for these indices. Column (1) regresses overplaying on the Tenacity and Diligence indices, and column (2) controls for plan of action, age, gender and degree (cf. columns (1) and (5) in Table 3). Columns (3) and (4) provide sample splits by degree and columns (5) and (6) by gender, analogously to Table 4. Column (7) checks for robustness to overplaying by more than 100 tokens, analogously to column (3) of Appendix Table 17. We find that the coefficient for the Tenacity Index is highly significant for all 7 columns, and its magnitude tends to be larger than the one obtained for the Grit Index in previous tables. The Diligence Index, however, is never significant, and its estimated coefficient remains close to zero for all columns. These findings jointly support the hypothesis that it is the tenacity component of grit that drives overplaying.[19]

[18] The correlation coefficient is 0.56 among our 138 participants.

[19] The results for tenacity and diligence are robust to all the specifications discussed in the previous subsection, although we do not present them all here. Furthermore, results are also robust to adding observations from 3

Table 7 decomposes the baseline regressions into different groups. A sample split by educational degree is presented in the first two columns and another sample split by gender in the last two columns. Throughout the four specifications, the estimated coefficient on the Tenacity Index is economically and statistically significant.

Graphically, the kernel density estimation for tenacity of those who overplayed and of those who have not shows a pronounced shift of the distribution (Figure 6). This is also corroborated in Figure 7 where the cumulative distribution of those who overplayed again first order stochastically dominates the distribution of those who have not. Instead, the analogous kernel density estimation for diligence shows very little difference (Figure 8). This provides additional visual assurance that the strong positive relationship between grit and overplaying is through tenacity.

We then ensure that our categorization is robust and not subjective in two different ways. First, we perform a confirmatory factor analysis (CFA) on the split of grit into tenacity and diligence. A CFA is a widely used tool that serves to confirm that a categorization into a number of factors, in this case tenacity and diligence, is robust (see Appendix D for details). The CFA is provided in the Appendix Table 22. Its indices provide empirical evidence of fit between the data and the model, thus serving as a validation of measurements of latent construct. Since our split of the Grit Index builds on a specific theory, the factorial validity is tested via a theory-driven CFA rather than a data-driven explanatory factor analysis. Following the standard procedure, we allow for the within-factor statistically significant error covariances (specifically, "I finish whatever I begin" with "New ideas and projects sometimes distract me from previous ones," "I finish whatever I begin" with "I have been obsessed with a certain idea or project for a short time but later lost interest," and "I am a hard worker" with "I am diligent").

Second, we consider alternative specifications of this split. To this end, we remove questions whose categorization carries any hint of ambiguity. Our results remain unchanged to these specifications. We do not discuss them all for the sake of brevity, but we provide here the most striking such alternative specification. This consists of removing all but the two clearest questions that measure tenacity ("Setbacks don't discourage me" and "I finish whatever I begin") and the two that measure diligence ("I am diligent" and "I am a hard worker"). Running all previous regressions on these indices leaves the significance levels unchanged and the coefficient in the same range (Appendix Table 23). Moreover, these four questions, which are split equally into two separate categories in our alternate reduced specification, are all grouped together in the standard classification under 'perseverance.' This further illustrates the different rationale behind classifying grit into tenacity and diligence compared to perseverance and consistency of interest.

Alternative decomposition into perseverance and consistency of interest

Our categorization is distinct from the standard one of 'perseverance' and 'consistency of interest.' Returning to the rowboat metaphor, a more perseverant person rows harder and does not relent when faced with obstacles, and one who has consistency of interest does not change his

sessions (48 observations) in which the Grit Questionnaire was provided before the actual games were played (see Appendix Table 21). For these additional sessions, the likelihood of overplaying remains unchanged and the distribution of the grit index is the same as for the previous sessions (Mann-Whitney Test).

mind often on his final destination. Neither perseverance nor consistency of interest, therefore, perfectly maps either to our diligence or tenacity category. Diligence only concerns working hard and not the reaction to obstacles faced, and therefore neither fully fits within perseverance nor consistency of interest. Tenacity focuses on refusal to steer away, not the strength of the rowing. It is also not obvious, from this image, whether either perseverance or consistency are natural candidate to explain overplaying. The difference in what these categories aim to capture is naturally reflected in the distinct partitioning. For instance, "I am diligent" and "Setbacks don't discourage me" both fall within perseverance in the standard classification. But the first falls within diligence, as it concerns the strength of rowing, while the second falls within tenacity, as it concerns the refusal to steer away.[20]

To compare the explanatory power of our categorization to the standard one, Table 8 replicates the same regressions as Table 6, but using the perseverance and consistency of interest split instead of our tenacity and diligence split. The coefficient estimates are clearly less stable. Furthermore, in some columns, perseverance is significant but consistency is not (e.g. column (3)), in some it is the reverse (e.g. column (4)), and in others neither is significant (e.g. columns (6) and (7)).

4 Further Results

4.1 Correlation between overplaying across games

As the order of the two games is always the same, one might be concerned about order effects. In particular, it might be that participants who got lucky in the first game overplay less in the second game. Table 9 looks at how winning in the first game is associated with overplaying in the second game. In particular, the new regressor of interest is a dummy variable taking value 1 if the subject overplayed in the first game. The dependent variable is still the usual dummy for overplaying in the second game.

In Column (1) we indeed find that overplaying in the first game predicts a greater likelihood of overplaying in the second game. Furthermore, this result still holds after controlling for the type of degree and the player's gender (Column (2)). In Column (3) we add the remaining controls and, importantly, also the Grit Index. While the dummy for winning in the first game is still statistically significant, the magnitude and statistical significance of the Grit Index is basically unaffected compared to our baseline result seen in the baseline table. This is consistent with the two variables having orthogonal variation to each other and indeed the correlation between them is only 0.037. In Column (4) we split the Grit Index into its two components: Tenacity Index and Diligence Index. A similar situation emerges: while the dummy for overplaying in the first game is still strongly significant, the effect of the Tenacity Index is almost unaltered compared to the baseline regression. The correlation between these two variables is again very small at -0.01. Finally, in the last two columns we remove the 30 subjects who won in the first

[20]To get a sense of the partial correlations across these variables, we first regress diligence on consistency and perseverance; we obtain an estimated coefficient of 0.37 for the former and 0.71 for the latter. We then set tenacity as the dependent variable, and obtain an estimated coefficient of 0.57 for consistency and 0.37 for perseverance. All four estimated coefficients are highly statistically significant.

game. Therefore, we are left with a sample size of 108 subjects. Reassuringly, the main findings of the paper are still economically and statistically significant.

4.2 Chasing losses

Prediction 3 states that subjects with a higher Tenacity Index are more likely to end with final earnings of either 0 or an amount greater or equal to the initial endowment of 2000 tokens. For this reason, an interpretation of the mechanism we are discussing is that tenacity provides a force towards chasing losses: the cost of failure incurred leads the agent to keep playing when losing rather than stopping.

Table 10 provides empirical support for this interpretation. The dependent variable is a dummy taking value 1 if the subject ended up with either 0 tokens or with more than 2000 tokens. The regressors of interest are the Tenacity Index and the Diligence Index. Once agents are in the loss domain (columns 2-4), i.e. have reaching a number of tokens below 2000, we find strong evidence for the fact that more tenacious individuals are more likely to end up with extreme outcomes, consistent with a chasing losses argument.

4.3 BDM valuations and Tenacity

We are cautious in interpreting the willingness to pay given by subjects in the BDM for their plan of action, as previously discussed, because of the complexity of the task. With this caveat, we analyze the relationship between the valuation given by subjects in the BDM and tenacity. We find that tenacity does not correlate with a higher valuation, which suggests that our results are not due to more tenacious subjects overplaying because of less attachment to their plan of action. It would be interesting to explore these points further in a future experiment.[21]

4.4 Procrastination, temptation and self-esteem

To confirm that there is an upside of grit present in this setting, we explore the link between grit (and its decomposition into tenacity and diligence) and self-reported procrastination, temptation and self-esteem. In particular, higher grit is typically associated with reporting lower procrastination and temptation problems and higher self-esteem.[22]

In Table 11 we regress our three outcome variables on grit (columns (1), (5) and (9)), and find estimated coefficients in line with our expectations. When using only tenacity (columns (2), (6), and (10)) or diligence (columns (3), (7), and (11)) as regressors, we still obtain that they are associated with lower procrastination and temptation, and at the same time more self-esteem.

Interestingly, when we regress procrastination on both tenacity and diligence together (columns (4), (8), and (12)), only the coefficient of diligence is highly statistically significant. Diligence has a clear upside, while the effect of tenacity is more ambiguous. For the case of temptation, for instance, there may well be an upside in keep focused on the task if it has already begun, but there is a temptation of sticking with the tempting activity if it is in motion as well. Overall,

[21] We thank an anonymous referee for this suggestion.

[22] These questions were asked to 118 of the 138 subjects.

the following pattern emerges: tenacity alone explains overplaying, while diligence has more explanatory power for lower temptation and procrastination and higher self-esteem.

4.5 Admitting having overplayed after the game

We close this section by discussing an additional survey question given to the subjects, near the end of the experiment. This question asks whether they have overplayed in the roulette or not. We then compare this answer to whether the agents did in fact overplay. In other words, we analyze whether agents 'admit' to having overplayed, by their own measure (Table 12). A reasonable conjecture is that while less stubborn (or tenacious) subjects would admit having overplayed, more stubborn would not. This is consistent with a view that tenacity also involves being stubborn in accepting having gone too far.

We first see in column (1) that the coefficient for overplaying is positive and highly significant, which confirms that overall, subjects do admit having overplayed. Interestingly, and consistent with the conjecture above, column (2) reveals that when tenacity is included, the coefficient for overplaying remains positive and highly significant, while the coefficient for tenacity is negative and significant. That is, more tenacious subjects refuse to admit right after the game that they did overplay. In columns (3) and (4) we add diligence and the plan of action to the regression, and find that neither of them is significant, while the coefficients for overplaying remains positive and highly significant, and the coefficient for tenacity remains negative and significant. Moreover, the same finding holds in column (5) when we condition the sample on overplayers. These results suggest that in addition to tenacity involving refusal to let go through incurring a cost of failure, it may also involve stubbornness in accepting having gone too far.

4.6 External Validity - Educational Outcomes

Here we consider how well diligence and tenacity explain educational outcomes in two different datasets. While much more detail is included in the Appendix, here we just briefly summarize the main findings.

The first dataset, discussed in Section C.1, consists of a survey on educational outcomes provided by the Interuniversity Consortium for Political and Social Research (ICPSR). Throughout Appendix Table 24, the Grit Index positively correlates with both the ACT test score and also Also, both tenacity and diligence are indicative of higher educational performance in various domains.

The second, discussed in Section C.2, consists of an online survey on grit, conscientiousness, and educational attainment (Appendix Table 25). Using conscientiousness as the dependent variable, we first confirm it is highly correlated with grit. When decomposing grit into tenacity and diligence, tenacity does not entirely capture conscientiousness. Consequently, understanding more precisely the degree to which our results extend to conscientiousness requires further exploration.

In Appendix Table 26 we document how educational outcomes are explained by the Conscientiousness Index and Grit Index, and the split into Diligence and Tenacity Indices. When regressing education on conscientiousness and diligence and tenacity, all three are significant without controls, but only diligence remains significant with controls as well.

These results confirm the upside of diligence discussed in our analysis (noting that diligence may have downsides too in some contexts) and shows that tenacity is more nuanced. Nonetheless, one should take these results mainly as indicative and suggestive, as a key element of our story (plan of action) is not observed.

5 Conclusions

Our results indicate that grittier subjects have a higher tendency to play past the point at which they would have liked to stop. We have further shown that when grit is split into tenacity and diligence, tenacity alone explains this tendency to overplay. Diligence, instead, explains lower procrastination and temptation problems within our experiment, and higher educational level when applied to an existing survey.

The upside of grit is well-known, but our analysis reveals that it may have an important flip side too. Individuals with higher grit may also have more difficulty in stopping and accepting failure, even when they would have liked to. This tendency is contained within the tenacity facet of grit, which itself has both the positive aspect associated with not giving up and the negative aspect associated with not letting go. Returning to the image of the rowboat introduced in the paper, resistance to steer away from the current route and onto another, which describes tenacity, intertwines stubbornness with steadfastness.

Caution is needed in interpreting our results as being conclusive that tenacity itself is the driver of grit. Future research should directly measure subjects' risk attitude to cleanly distinguish the role of risk from that of tenacity. Moreover, there may be other factors that correlate with tenacity that we do not observe in this experiment- for instance, it could be that more tenacious subjects are more likely to find the roulette game more enjoyable than they had anticipated, which we do not measure or observe. Here too, future research is required to better understand whether another interpretation than the one offered in this paper can be ruled out or not.

Our findings raise new questions. Under the interpretation offered in this model, the upside of tenacity may well outweigh its downside in many contexts, but in others it may prove costly. There may also be other potential downsides of grit; Duckworth and Eskreis-Winkler (2015), for instance, speculate that those who are more likely to stay the course may also ignore new opportunities. This tension should be investigated further to shed light on addressing the flip side of grit and tenacity without diminishing the positive side.

Table 1: Descriptive statistics

Note: Overplaying is a dummy variable taking value 1 if the subject played beyond his plan of action. Plan of Action is the log of 1 plus the planned minimum bound, where the planned minimum bound ranges between 0 and 2000 tokens. The Grit, Tenacity, and Diligence Indices range from 1 to 5. The Procrastination, Temptation and Self-esteem measures are self-reported measures that also range from 1 to 5.

Variable	Obs.	Mean	Std. Dev.	Min	Max	Potential Min	Potential Max
Overplaying (Dummy)	138	.35	.48	0	1	0	1
Overplaying (Rounds)	138	6.637	17.39	0	131		
Grit Index	138	3.38	.54	2	4.5	0	5
Plan of Action (level)	138	877.54	578.19	0	2000	0	2000
Plan of Action (log)	138	6.21	1.73	0	7.6	0	7.6
Tenacity Index	138	3.22	.57	2	4.6	0	5
Diligence Index	138	3.65	.68	2	5	0	5
Locus Index	138	.47	.17	.06	.88	0	1
Age	138	21.72	3.73	18	47		
Female	138	.58	.49	0	1	0	1
Technical Degree	138	.51	.50	0	1	0	1
Self-esteem	138	3.14	.99	1	5	1	5
Procrastination	118	3.29	.80	1	5	1	5
Temptation	118	3.31	.85	1	5	1	5

Table 2: Descriptive statistics - Correlations with Grit, Tenacity, and Diligence

Note: Here we take the same variables previously defined in Table 1 and show the correlation coefficients with respect to: (i) Grit Index; (ii) Tenacity Index; (iii) Diligence Index.

Variable	Observations	Correlation with:		
		Grit Index	Tenacity Index	Diligence Index
Overplaying (Dummy)	138	.29	.33	.16
Plan of Action (log)	138	.06	.07	.02
Locus Index	138	-.19	-.19	-.14
Age	138	.16	.16	.11
Female	138	.07	.07	.04
Technical Degree	138	.08	.00	.17
Self-esteem	138	.32	.24	.33
Procrastination	118	-.45	-.35	-.45
Temptation	118	-.36	-.30	-.34

Table 3: Baseline regressions

Note: An Ordinary Least Squares specification is used. The dependent variable is a dummy variable taking the value 1 in case the subject overplayed. Grit Index takes values between 1 and 5 and increases with the level of grit. Plan of Action is the log of 1 plus the planned minimum bound. Standard errors are robust to heteroskedasticity. Significance: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Dep.Var.: Overplaying	(1)	(2)	(3)	(4)	(5)
Grit Index	0.259***	0.248***	0.237***	0.248***	0.260***
	(0.072)	(0.070)	(0.071)	(0.070)	(0.071)
Plan of Action		0.067***	0.068***	0.072***	0.074***
		(0.013)	(0.013)	(0.014)	(0.014)
ln(Age)			0.236	0.154	0.076
			(0.274)	(0.267)	(0.262)
D(Female)				-0.124	-0.156*
				(0.080)	(0.082)
Technical Degree					-0.095
					(0.080)
Observations	138	138	138	138	138
R-squared	0.087	0.145	0.150	0.165	0.174

Table 4: Sample split by degree and gender

Note: This table splits the full sample based on the type of undergraduate degree studied or the gender of the subject. Column (1) uses the subset of individuals who studied a technical degree, while column (2) uses the remaining subjects. Column (3) uses only data on women, while column (4) only men. Significance: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

	(1)	(2)	(3)	(4)
Dep. Var.: Overplaying	Technical	Non-technical	Female	Male
Grit Index	0.206*	0.287***	0.312***	0.201*
	(0.106)	(0.093)	(0.094)	(0.104)
Plan of Action	0.064***	0.086***	0.063**	0.073***
	(0.014)	(0.026)	(0.028)	(0.015)
ln(Age)	-0.468	0.583**	-0.369	0.573**
	(0.402)	(0.270)	(0.372)	(0.259)
Observations	71	67	80	58
R-squared	0.117	0.242	0.157	0.205

Table 5: Decomposition of Grit Index into tenacity and diligence

Note: The table below decomposes the 8 questions of the Grit Index and allocates each question into a cell. The rows are defined by our proposed split of the index: Tenacity and Diligence. The columns are defined by the split suggested by Duckworth and Quinn (2009): Perseverance and Consistency. Furthermore, the four questions in italics are the ones kept in our main robustness indices of tenacity and diligence.

	Perseverance	**Consistency**
Tenacity	*Setbacks don't discourage me.* *I finish whatever I begin.*	New ideas and projects sometimes distract me from previous ones. I often set a goal but later choose to pursue a different one. I have been obsessed with a certain idea or project for a short time but later lost interest.
Diligence	*I am diligent.* *I am a hard worker.*	I have difficulty maintaining my focus on projects that take more than a few months to complete.

Table 6: Splitting Grit Index into tenacity and diligence

Note: This table builds on Table 3, but now splits Grit Index into Tenacity Index and Diligence Index. Column (1) uses the full sample of observations with no controls, similar to column (1) on Table 3. Column (2) adds the full battery of control variables, as in column (5) on Table 3. Columns (3) to (6) split the data by educational degree and gender, similar to Table 10. Finally, column (7) drops the 10 subjects who overplayed by less than 100 tokens. Significance: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Dep. Var.: Overplaying	(1)	(2)	(3) Technical	(4) Non-technical	(5) Female	(6) Male	(7) >100
Tenacity Index	0.298***	0.278***	0.237**	0.371***	0.280***	0.313***	0.227***
	(0.074)	(0.068)	(0.091)	(0.096)	(0.095)	(0.096)	(0.073)
Diligence Index	-0.028	-0.011	-0.033	-0.058	0.042	-0.107	-0.017
	(0.073)	(0.071)	(0.108)	(0.098)	(0.092)	(0.104)	(0.075)
Plan of Action		0.071***	0.057***	0.093***	0.068**	0.063***	0.077***
		(0.013)	(0.015)	(0.030)	(0.030)	(0.018)	(0.013)
ln(Age)		0.086	-0.494	0.637**	-0.357	0.519*	0.104
		(0.274)	(0.437)	(0.261)	(0.390)	(0.277)	(0.265)
D(Female)		-0.150*					-0.180**
		(0.081)					(0.082)
Technical Degree		-0.068					-0.088
		(0.083)					(0.084)
Observations	138	138	71	67	80	58	128
R-squared	0.112	0.192	0.136	0.280	0.167	0.250	0.186

Table 7: Sample split and Tenacity Index

Note: This table splits the full sample based on the type of undergraduate degree studied or the gender of the subject. Column (1) uses the subset of individuals who studied a technical degree, while column (2) uses the remaining subjects. Column (3) uses only data on women, while column (4) only men. The regressors of interest are the Tenacity Index and the Diligence Index. Significance: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Dep. Var.: Overplaying	(1) Technical	(2) Non-technical	(3) Female	(4) Male
Tenacity Index	0.237**	0.371***	0.280***	0.313***
	(0.091)	(0.096)	(0.095)	(0.096)
Diligence Index	-0.033	-0.058	0.042	-0.107
	(0.108)	(0.098)	(0.092)	(0.104)
Plan of Action	0.057***	0.093***	0.068**	0.063***
	(0.015)	(0.030)	(0.030)	(0.018)
ln(Age)	-0.494	0.637**	-0.357	0.519*
	(0.437)	(0.261)	(0.390)	(0.277)
Constant	0.846	-3.166***	-0.099	-2.197**
	(1.350)	(0.807)	(1.180)	(0.910)
Observations	71	67	80	58
R-squared	0.136	0.280	0.167	0.250

Table 8: Splitting Grit Index to consistency and perseverance

Note: This table builds on Table 3, but now splits Grit Index into Consistency Index and Perseverance Index. Column (1) uses the full sample of observations with no controls, similar to column (1) on Table 3. Column (2) adds the full battery of control variables, as in column (5) on Table 3. Columns (3) to (6) split the data by educational degree and gender, similar to Table 10. Finally, column (7) drops the 10 subjects who overplayed by less than 100 tokens. Significance: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Dep. Var.: Overplaying	(1)	(2)	(3) Technical	(4) Non-technical	(5) Female	(6) Male	(7) >100
Consistency Index	0.136**	0.104	-0.003	0.185**	0.107	0.107	0.092
	(0.066)	(0.063)	(0.093)	(0.080)	(0.083)	(0.102)	(0.064)
Perseverence Index	0.122*	0.162**	0.242**	0.096	0.215**	0.093	0.114
	(0.072)	(0.070)	(0.109)	(0.093)	(0.100)	(0.093)	(0.075)
Plan of Action		0.076***	0.078***	0.086***	0.062**	0.073***	0.079***
		(0.015)	(0.020)	(0.028)	(0.027)	(0.017)	(0.015)
ln(Age)		0.072	-0.510	0.588**	-0.396	0.569**	0.110
		(0.267)	(0.395)	(0.271)	(0.378)	(0.270)	(0.260)
D(Female)		-0.157*					-0.185**
		(0.083)					(0.084)
Technical Degree		-0.105					-0.119
		(0.082)					(0.085)
Observations	138	138	71	67	80	58	128
R-squared	0.087	0.176	0.144	0.247	0.163	0.205	0.172

Table 9: Effect of Winning in the First Game on Overplaying

Note: This table builds on Table 3 and Table 6, but now adds a regressor "Dummy Winner in the First Game". All specifications use the full sample, except the last two columns where we drop the 30 subjects who won in the First Game. Significance: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Dep. Var.: Overplaying in the Second Game	(1)	(2)	(3)	(4)	(5)	(6)
Grit Index			0.254***		0.258***	
			(0.070)		(0.076)	
Tenacity Index				0.309***		0.298***
				(0.064)		(0.068)
Diligence Index				-0.046		-0.034
				(0.065)		(0.074)
Dummy Winner in First Game	0.280***	0.279***	0.316***	0.345***		
	(0.101)	(0.102)	(0.094)	(0.092)		
Technical Degree		-0.026	-0.083	-0.047	-0.044	-0.016
		(0.085)	(0.077)	(0.078)	(0.088)	(0.087)
D(Female)		-0.096	-0.169**	-0.163**	-0.176**	-0.181**
		(0.087)	(0.079)	(0.077)	(0.088)	(0.086)
Plan of Action			0.084***	0.082***	0.072***	0.071***
			(0.016)	(0.015)	(0.016)	(0.015)
ln(Age)			-0.051	-0.050	-0.353	-0.325
			(0.232)	(0.245)	(0.301)	(0.318)
Constant	0.287***	0.356***	-0.804	-0.788	0.165	0.107
	(0.044)	(0.092)	(0.698)	(0.748)	(0.892)	(0.962)
Observations	138	138	138	138	108	108
R-squared	0.059	0.068	0.245	0.275	0.186	0.214

Table 10: Effect of Tenacity on Extreme Outcomes

Dep.Var.: Finish Second Game with 0 or $\geq$ 2000	(1)	(2)	(3)	(4)
Tenacity Index	0.132	0.217**	0.195*	0.355***
	(0.096)	(0.100)	(0.115)	(0.123)
Diligence Index	-0.030	-0.075	0.027	-0.031
	(0.085)	(0.088)	(0.100)	(0.115)
Technical Degree	-0.044	-0.108	-0.098	-0.015
	(0.098)	(0.100)	(0.112)	(0.158)
D(Female)	-0.118	-0.149	-0.230**	-0.166
	(0.098)	(0.101)	(0.114)	(0.140)
ln(Age)	0.182	0.174	0.244	0.700**
	(0.298)	(0.294)	(0.305)	(0.324)
Plan of Action	-0.022	-0.045*	-0.039	-0.038
	(0.026)	(0.026)	(0.025)	(0.030)
Constant	-0.192	-0.103	-0.648	-2.351**
	(0.963)	(0.953)	(0.959)	(1.029)
Minimum Tokens Reached	<2000	<1900	<1500	<1000
Observations	129	116	85	52
R-squared	0.042	0.099	0.148	0.305

Table 11: Post-Questions on procrastination, temptation, and self-esteem

Note: In columns (1) to (4), the dependent variable is the degree of procrastination problems, in an increasing range from 1 to 5. Columns (5) to (8) rather use the degree of temptation problems (also in an increasing range from 1 to 5) as dependent variable. Finally, the last four columns look at self-esteem as an outcome variable. Significance: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Dep.Var.:	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
	Procrastination				Temptation				Self-esteem			
Grit Index	-0.660***				-0.570***				-0.576***			
	(0.141)				(0.130)				(0.156)			
Tenacity Index		-0.497***		-0.217		-0.454***		-0.249		0.425***		0.150
		(0.150)		(0.184)		(0.131)		(0.162)		(0.156)		(0.177)
Diligence Index			-0.525***	-0.427***			-0.425***	-0.311**			0.480***	0.409***
			(0.096)	(0.128)			(0.110)	(0.136)			(0.128)	(0.150)
Observations	118	118	118	118	118	118	118	118	138	138	138	138
R-squared	0.198	0.125	0.201	0.218	0.129	0.091	0.115	0.134	0.099	0.060	0.108	0.113

Table 12: The Effect of Stubbornness on Admitting Overplaying

Note: In this table the dependent variable is based on a question posed to the students after the games ended. It is a dummy taking value 1 if the student claims to have overplayed during the Roulette game. Column (1) includes the actual Overplaying variable already used in previous tables. Column (2) adds the Tenacity Index and column (3) also includes the Diligence Index. Column (4) also adds the Plan of Action and, finally, column (5) limits the sample to the 56 students who actually overplayed in the second game. Significance: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Dep. Var.: Admit Overplaying	(1)	(2)	(3)	(4)	(5)
Overplaying	0.331***	0.373***	0.372***	0.401***	
	(0.082)	(0.079)	(0.080)	(0.082)	
Tenacity Index		-0.152**	-0.196**	-0.193**	-0.359***
		(0.069)	(0.086)	(0.086)	(0.121)
Diligence Index			0.065	0.062	0.128
			(0.069)	(0.069)	(0.098)
Plan of Action				-0.033	0.022
				(0.025)	(0.073)
Constant	0.366***	0.838***	0.741***	0.933***	1.274**
	(0.054)	(0.225)	(0.240)	(0.292)	(0.623)
Observations	138	138	138	138	56
R-squared	0.105	0.134	0.139	0.151	0.141

Figure 1: Roulette Pictures

Note: Snapshots of the roulette game. In the first picture, the subject places a bet (at the start of the game) of 50 tokens on the number 4, 100 tokens on the number 10 and 50 tokens on the color red. The second picture takes place later in the game, and the subject sees the history of the last five outcomes (with the last being 12 red), how much he has won and bet following the previous spin, and his remaining tokens. He has not yet placed the new bet. He is always free to quit, and he is allowed to spin even without placing any bet.

Figure 2: Histogram of the Plan of Action

Note: The histogram shows the number of rounds a subjects plan to play, i.e. their plan of action.

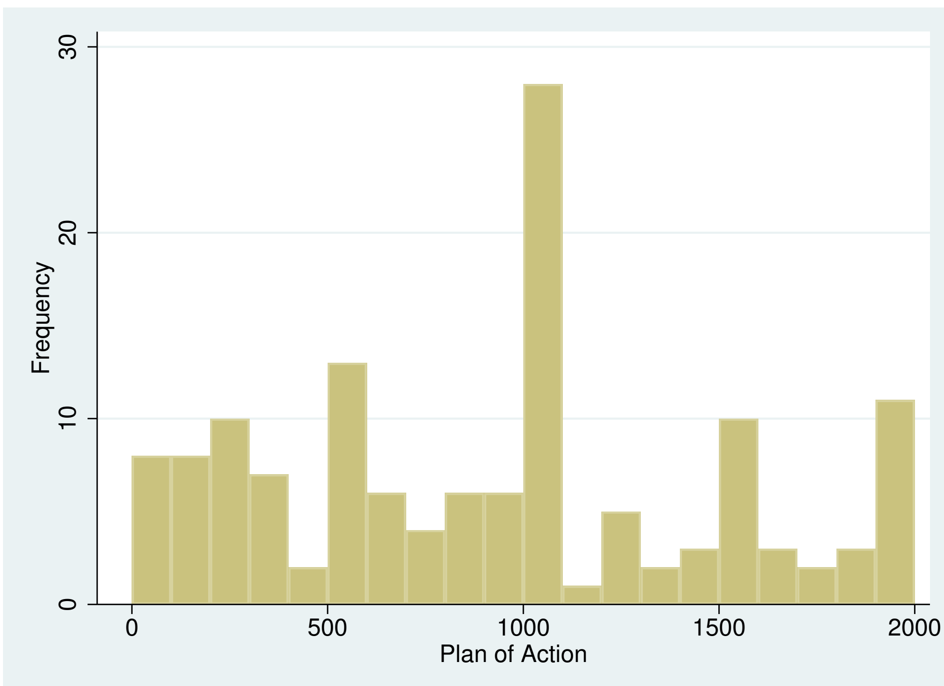
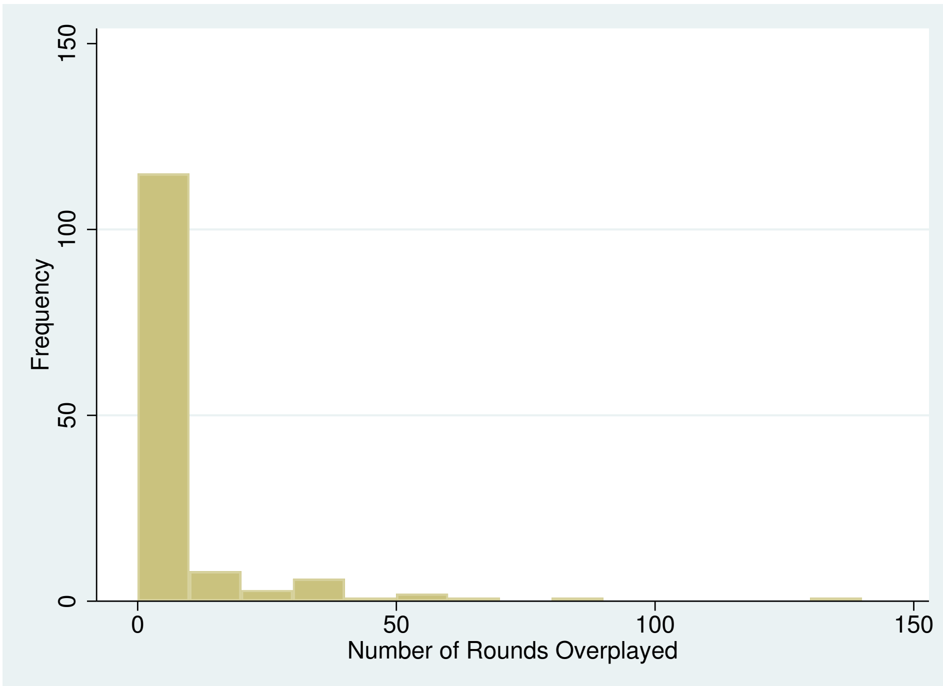


Figure 3: Histogram of Number of Rounds Overplayed

Note: These histograms show the number of rounds a subjects overplays, i.e. goes beyond the plan of action.

Panel A: Full sample of 138 subjects



Panel B: Conditional on 48 subjects who overplayed

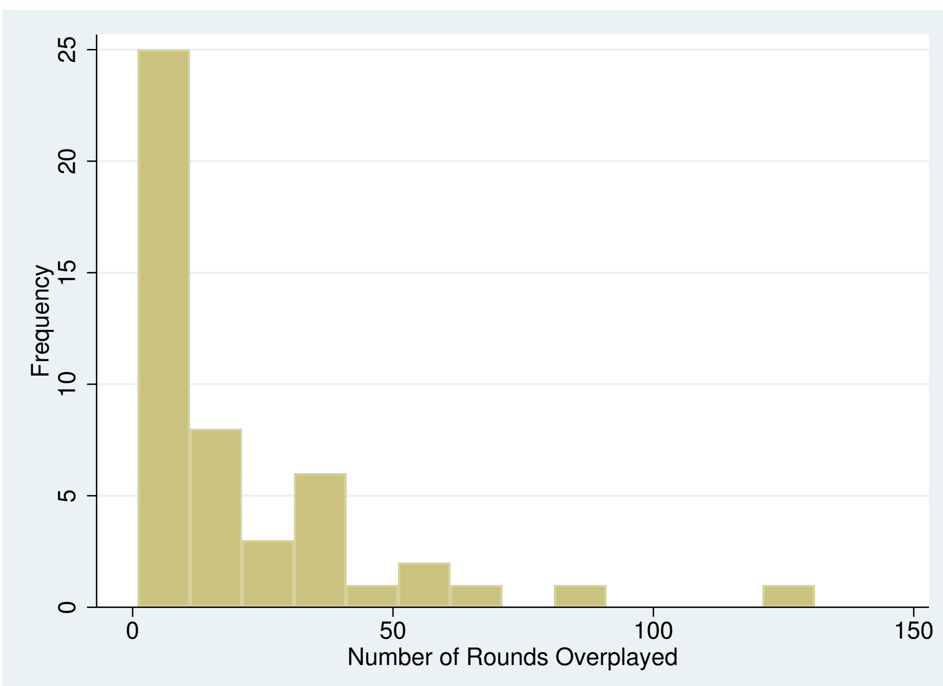


Figure 4: Kernel Density - Grit Index

Note: This figure separately traces the kernel density distributions for the Grit Index for subjects who did not overplay (blue dashed line) and for those who overplayed (red solid line).

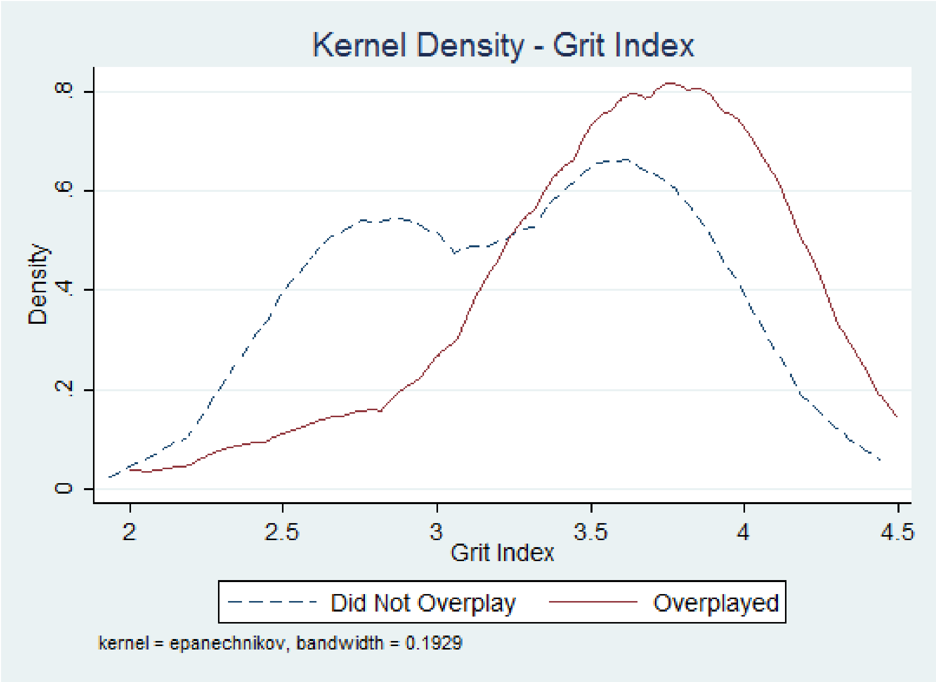


Figure 5: Cumulative Density Function - Grit Index

Note: Building on Figure 4, this figure separately traces the cumulated density functions for subjects who did not overplay (blue dashed line) and for those who overplayed (red solid line).

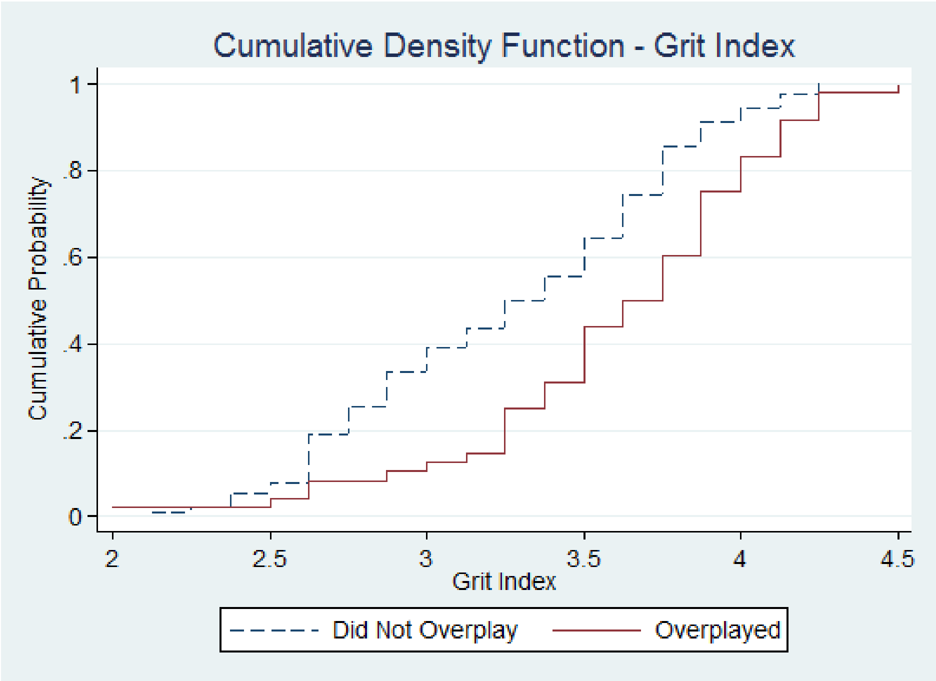


Figure 6: Kernel Density - Tenacity Index

Note: This figure separately traces the kernel density distributions for the Tenacity Index for subjects who did not overplay (blue dashed line) and for those who overplayed (red solid line).

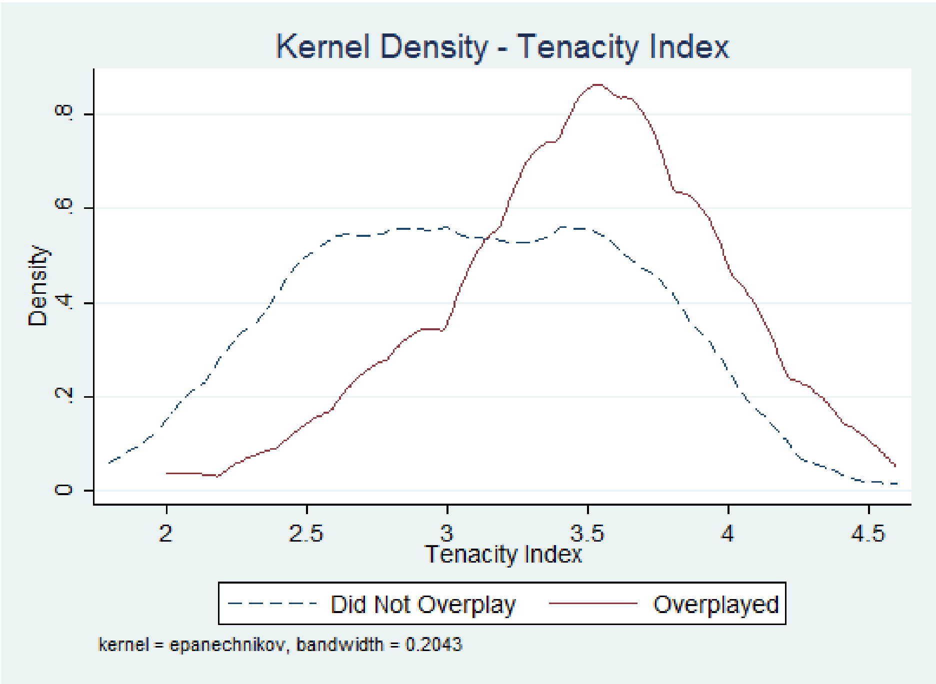


Figure 7: Cumulative Density Function - Tenacity Index

Note: Building on Figure 6, this figure separately traces the cumulated density functions for subjects who did not overplay (blue dashed line) and for those who overplayed (red solid line).

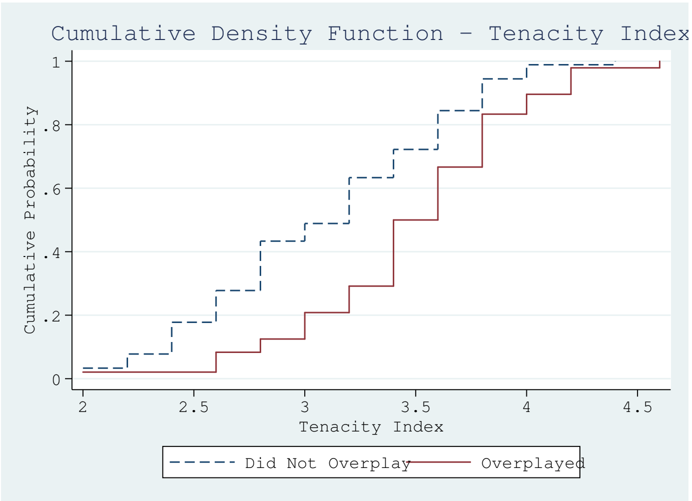
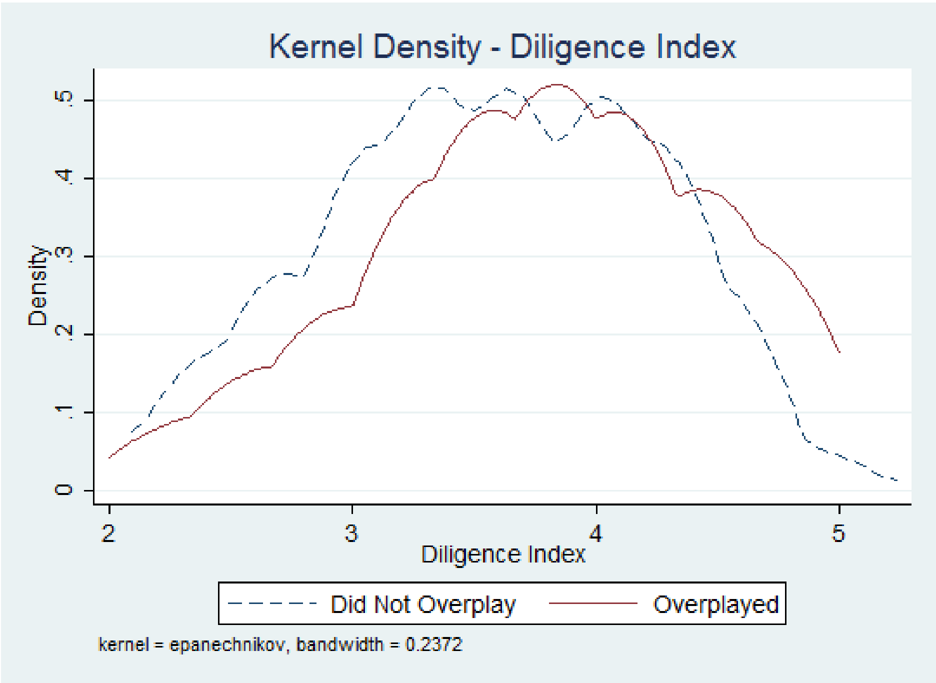


Figure 8: Kernel Density - Diligence Index

Note: This figure separately traces the kernel density distributions for the Diligence Index for subjects who did not overplay (blue dashed line) and for those who overplayed (red solid line).



References

[1] Achtziger, A., Bayer, U. C., and Gollwitzer, P. M. Committing to implementation intentions: Attention and memory effects for selected situational cues. *Motivation and Emotion 36* (2012), 287–300.

[2] Alan, S., Boneva, T., and Ertac, S. Ever failed, try again, succeed better: Results from a randomized educational intervention on grit. *Quarterly Journal of Economics 134*, 3 (2019), 1121–1162.

[3] Alan, S., and Ertac, S. Good things come to those who (are taught how to) wait: Results from a randomized educational intervention on time preference. *Working Paper* (2014).

[4] Alan, S., and Ertac, S. Patience, self-control and the demand for commitment: Evidence from a large-scale field experiment. *Journal of Economic Behavior and Organization 115* (2015), 111–122.

[5] Almlund, M., Duckworth, A. L., Heckman, J., and Kautz, T. Personality psychology and economics. *Handbook of the Economics of Education 4 (Chapter 1)* (2011).

[6] Amador, M., Werning, I., and Angeletos, G.-M. Commitment vs. flexibility. *Econometrica 74*, 2 (2006), 365–396.

[7] Ameriks, J., Caplin, A., Leahy, J., and Tyler, T. Measuring self-control problems. *American Economic Review 97*, 3 (2007), 966–972.

[8] Andrade, E. B., and Iyer, G. Planned versus actual betting in sequential gambles. *Journal of Marketing Research 46*, 3 (2009), 372–383.

[9] Ashraf, N., Karlan, D., and Yin, W. Tying Odysseus to the mast: Evidence from a commitment savings product in the philippines. *Quarterly Journal of Economics 121*, 2 (2006), 635–672.

[10] Augenblick, N., Niederle, M., and Sprenger, C. Working over time: dynamic inconsistency in real effort tasks. *Quarterly Journal of Economics* (2015), 1067–1115.

[11] Barberis, N. A model of casino gambling. *Management Science 58*, 1 (2012), 35–51.

[12] Barbosa, S. D., Gerhardt, M. W., and Kickul, J. R. The role of cognitive style and risk preference on entrepreneurial self-efficacy and entrepreneurial intentions. *Journal of Leadership & Organizational Studies 13*, 4 (2007), 86–104.

[13] Borghans, L., Duckworth, A. L., Heckman, J. J., and ter Weel, B. The economics and psychology of personality traits. *Journal of Human Resources 43*, 4 (2008), 972–1059s.

[14] Boyce, C. J., Wood, A. M., and Brown, G. D. The dark side of conscientiousness: Consciencious people experience greater drops in life satisfaction following unemployment. *Journal of Research in Personality 44* (2010), 535–539.

[15] Burks, S. V., Lewis, C., Kivi, P. A., Wiener, A., Anderson, J. E., Gotte, L., DeYoung, C. G., and Rustichini, A. Cognitive skills, personality, and economic preferences in collegiate success. *Journal of Economic Behavior & Organization 115* (2015), 30–44.

[16] Crede, M., Tynan, M., and Harms, P. Much ado about grit: A meta-analytic synthesis of the grit literature. *Journal of Personality and Social Psychology* (2016).

[17] Cubel, M., Nuevo-Chiquero, A., Sanchez-Pages, S., and Vidal-Fernandez, M. Do personality traits affect productivity? Evidence from the lab. *Economic Journal forthcoming* (2016).

[18] Cunha, F., Heckman, J. J., and Schennach, S. M. Estimating the technology of cognitive and noncognitive skill formation. *Econometrica 78*, 3 (2010), 883–931.

[19] Dobbie, W., and Fryer Jr., R. G. The medium-term impacts of high-achieving charter schools. *Journal of Political Economy 123*, 5 (2015), 985–1037.

[20] Duckworth, A., and Eskreis-Winkler, L. Grit. In *International Encyclopedia of the Social and Behavioral Sciences. 2nd edition*, J. D. Wright, Ed. Elsevier, 2015.

[21] Duckworth, A. L., and Gross, J. J. Self-control and grit: related but separable determinants of success. *Association for Psychological Science 23*, 5 (2014), 319–325.

[22] Duckworth, A. L., Peterson, C., Matthews, M. D., and Kelly, D. R. Grit: Perseverance and passion for long-term goals. *Journal of Personality and Social Psychology 92*, 6 (2007), 1087–1101.

[23] Duckworth, A. L., and Quinn, P. D. Development and validation of the short grit scale (Grit-S). *Journal of Personality Assessment 91*, 2 (2009), 166–174.

[24] Duckworth, A. L., and Seligman, M. E. Self-discipline outdoes IQ in predicting academic performance for adolescents. *Psychological Science 16*, 12 (2005), 939–944.

[25] Ebert, S., and Strack, P. Until the bitter end: On prospect theory in a dynamic context. *American Economic Review 105*, 4 (2015), 1618–1633.

[26] Ebert, S., and Strack, P. Never, ever getting started: On prospect theory without commitment. *Mimeo* (2018).

[27] Fischbacher, U., Hoffmann, G., and Schudy, S. The causal effect of stop-loss and take-gain orders on the disposition effect. *Review of Financial Studies 30*, 6 (2017), 2010–2129.

[28] Galasso, A., and Simcoe, T. CEO overconfidence and innovation. *Management Science 57* (2011), 1469–1484.

[29] Gerhards, L., and Gravert, C. Grit trumps talent? An experimental approach. *Mimeo* (2017).

[30] Gill, D., and Prowse, V. Cognitive ability, character skills, and learning to play equilibrium: A level-k analysis. *Journal of Political Economy 126*, 4 (2016), 1619–1676.

[31] Gneezy, U. Deception: the role of consequences. *American Economic Review 95*, 1 (2005), 384–394.

[32] Gollwitzer, P. M. Implementation intentions: Strong effects of simple plans. *American Psychologist 54* (1999), 493–503.

[33] Gul, F., and Pesendorfer, W. Temptation and self-control. *Econometrica 69*, 6 (2001), 1403–1435.

[34] Heckman, J. J., and Rubinstein, Y. The importance of noncognitive skills: Lessons from the GED testing program. *American Economic Review 91*, 2 (2001), 145–149.

[35] Heimer, R., Iliewa, Z., Imas, A., and Weber, M. Dynamic inconsistency in risky choice: Evidence from the lab and field. *Mimeo* (2020).

[36] Heineck, G. Does it pay to be nice? personality and earnings in the uk. *ILR Review 64*, 5 (2011), 1020–1038.

[37] Kaur, S., Kremer, M., and Mullainathan, S. Self-control at work. *Journal of Political Economy 123*, 6 (2015), 1227–1277.

[38] Kautz, T., Heckman, J., Diris, R., ter Weel, B., and Borghans, L. Fostering and measuring skills: Improving cognitive and non-cognitive skills to promote lifetime success. *NBER Working Paper 20749* (2014).

[39] Laibson, D. Golden eggs and hyperbolic discounting. *Quarterly Journal of Economics 112*, 2 (1997), 443–477.

[40] Lucas, G., Gratch, J., Cheng, L., and Marsella, S. When the going gets tough: Grit predicts costly perseverance. *Journal of Research in Personality 59* (2015), 15–22.

[41] MaCann, C., Duckworth, A., and Roberts, R. Empirical identification of the major facets of conscientiousness. *Learning and Individual Differences 19* (2009), 451–458.

[42] Malmendier, U., and Tate, G. CEO overconfidence and corporate investment. *Journal of Finance 60*, 6 (2005), 2661–2700.

[43] Odean, T. Are investors reluctant to realize their losses? *Journal of Finance 53* (1998), 1775–1798.

[44] O'Donoghue, T., and Rabin, M. Doing it now or later. *American Economic Review 89*, 1 (1999), 103–124.

[45] O'Donoghue, T., and Rabin, M. Present bias: Lessons learned and to be learned. *American Economic Review 105*, 5 (2015), 273–279.

[46] Papageorge, N., Ronda, V., and Zheng, Y. The economic value of breaking bad: Misbehavior, schooling and the labor market. *Mimeo* (2017).

[47] Ploner, M. Hold on to it? An experimental analysis of the disposition effect. *Judgment and Decision Making 12*, 2 (2017), 118–127.

[48] Proto, E., Rustichini, A., and Sofianos, A. Intelligence, personality and gains from cooperation in repeated interactions. *Journal of Political Economy Forthcoming* (2018).

[49] Roberts, B. W., Kuncel, N. R., Shiner, R., Caspi, A., and Goldberg, L. R. The power of personality: The comparative validity of personality traits, socioeconomic status, and cognitive ability for predicting important life outcomes. *Perspectives in Psychological Sciences 2* (2007), 313–345.

[50] Rotter, J. B. *Social learning and clinical psychology*. Prentice-Hall, Englewood Cliffs, NJ, 1954.

[51] Rueter, A., Abram, S., III, A. M., and Rustichini, A. The goal priority network as a neural substrate of conscientiousness,. *Human Brain Mapping Forthcoming* (2018).

[52] Sapienza, P., Zingales, L., and Maestripieri, D. Gender differences in financial risk aversion and career choices are affected by testosterone. *Proceedings of the National Academy of Sciences 106*, 36 (2009), 15268–15273.

[53] Shefrin, H., and Statman, M. The disposition to sell winners too early and ride losers too long: Theory and evidence. *Journal of Finance 40*, 3 (1985), 777–790.

[54] Sprenger, C. Judging experimental evidence on dynamic inconsistency. *American Economic Review: Papers & Proceedings 105*, 5 (2015), 280–285.

[55] Strotz, R. H. Myopia and inconsistency in dynamic utility maximization. *Review of Economics Studies 23*, 3 (1956).

Appendix (For Online Publication)

A Logistics of the Experiment

The experiment was conducted at the Behavioral Sciences Laboratory (BESLab) of Pompeu Fabra University (UPF), over 8 sessions, and included 138 subjects. Subjects were students of UPF, recruited using the BESLab system. No subject took part in more than one session. Subjects were paid 3 euros for showing up. They were also paid based on their earnings in one of the games (chosen at random) plus 4 euros for the end of the experiment survey. Earnings ranged from 7 euros to 60 euros. All earnings in the game were in tokens, where 150 tokens are worth 1 euro. Experimental sessions typically ranged between 60 and 75 minutes, with some subjects playing for longer (no time constraint was imposed).

A.1 Instructions of the Experiment

The main sequence of the experiment is as follows. The subjects were informed that they would be paid at random between the games, that all the games were single-agent, and that there were no right or wrong answers. In particular their first (Qualtrics) screens had the following information (translated from Spanish):

***

Welcome to the experiment. Please, do not use mobile phones or any other electronic devices.
Talking to other participants is not allowed. Please raise your hand if you have any questions.
Important: Do not close the web browser at any time!
During the experiment, your earnings in each game will be counted in tokens. A euro is worth 150 tokens. At the end of the experiment, you will be paid in euros.
Your earnings in each game will be independent of each other. At the end you will randomly be paid based on the earnings of one of these games. For example, if in a game you end up with 100 tokens and in another one you end up with 200 tokens, then your payment will be either 100 or 200 tokens, chosen at random.
In this experiment there are no right or wrong questions. All answers will only depend on your preferences, and different people will answer the questions in different ways. Simply answer the questions based on your preferences.
***

The first phase of the game consists of eliciting the subjects' plan of actions. We ask them to state their preferences for the first game, which we use as a robustness check for the plan of action. This is done as follows:

***

Consider the next game:
In each round of the game, you will choose a number between 1 and 38, and the computer will then also randomly choose a number between 1 and 38. Each number has the same likelihood of being chosen. If your chosen number is the same as the computer's, then you win 3200 tokens. If not, then you lose 100 tokens. If you win a round, then you cannot continue playing. That is, at most you can only win once.
If you start with 2000 tokens, after how many rounds would you like to stop playing?

***

They can choose any number between "0 (and you will end up with 2000 tokens)" and "20 (and if you don't win you will end up with 0 tokens)."
We then asked them for their plan of action of the roulette game. We explained the rules, but subjects were familiar with this game.

***
Consider this second game. The computer will choose a number between 1 and 36, as well as 2 additional numbers that we will call 0 and 00. Each number has the same likelihood of being chosen. In addition, half of the numbers are red and the other half are black.
You can choose a number between 1 and 36. You can also choose a color (red/black).
If your chosen number is the same as the computer's, then you get back your bet and you also receive 35 extra tokens for each token you bet. If your number is not selected, then you lose your bet and do not receive anything.
Similarly, if you have chosen a color and it is selected, then you get back your bet and additionally receive 1 extra token for each token you bet. If your color is not selected, then you lose your bet and do not receive anything.
You can play for as long as you want, provided you have enough tokens. There is a maximum of 8000 tokens that you can win. You can stop playing at any time. For this game, the game does not end if your chosen number is the same as the computer's.
Example: Suppose that you bet 100 tokens on number 15.
If the computer also chooses number 15, then you receive 3600 tokens (the 100 tokens of your initial bet and 3500 additional tokens).
If the computer chooses either 0, 00, or any other number between 1 and 36 that is not number 15, then you lose the 100 tokens you bet.
Now, suppose you bet 200 tokens on the color red. If the computer also chooses red, then you receive 400 tokens (the 200 tokens of your initial bet and 200 additional tokens). If not, then you lose the 200 tokens you bet.
If you start with 2000 tokens, what is the range of tokens inside which you would like to keep playing? [They enter the minimum and maximum tokens.]
***

They are given the choice between 0 and 2000 tokens, in increments of 100. We also used a simplified Becker-Degroot-Marshak (BDM) mechanism, although here they have no incentive to lie. In particular, we first explain in detail the mechanism at the beginning of the experiment. Then, this simplified mechanism is used only to ensure that they have incentives to truthfully report the stated preferences in the main game. (We rely on the stated preferences because the complexity of this setting would make the provided willingness to trade amount overly noisy, and these amounts are not useful for our objectives.) As noted previously, in the first game, designed as a robustness check, a high percentage of subjects play consistently with their plan of action, confirming that the stated preferences are reliable.
The mechanism used is as follows. After game 1, once the subjects have chosen their plan of action in game 1, they are then asked the following. "Consider the case in which you are given one of the following two options. Option 1: You are given 2000 tokens and you would have to play exactly 20 rounds (unless you win before). Option 2: You are given 2000 tokens and you would play according to the preferences you indicated, namely [quantity chosen by the subject] (unless you win before). How many tokens would you need to receive to prefer option 1? Remember

that the trading mechanism is always the one that you have seen at the start of the experiment, and that the best is to choose the amount that corresponds to your preferences." After game 2, the analogous question is asked for the setting: "Consider the case in which you are given one of the following two options. Option 1: You are given 2000 tokens and you would have to play until you lose all your tokens or win the maximum amount (8000 tokens). Option 2: You are given 2000 tokens and you will play according to the range that you have just indicated. How many tokens would you need to receive to prefer option 1? Remember that the trading mechanism is always the one that you have seen at the start of the experiment, and that the best is to choose the amount that corresponds to your preferences." Notice that subjects should anticipate the question when providing their plan of action since it follows the question of game 1, and that it is incentive compatible for them to give the correct plan beforehand. More precisely, it is incentive compatible both to give the preferred plan of action, and then the correct willingness to accept (for option 1). As mentioned above, however, from the complexity of the question we do not think that the willingness to trade is reliable, but it is not relevant for our purposes.

Before moving on to the second phase of the experiment (the actual games), we ask a variety of additional questions to reduce the salience of the previous ones. The subjects are informed that these questions will not matter for their earnings (this is done so that subjects do not incur much cognitive strain for these questions):

***

Before proceeding to play the games on which your earnings will be based, please answer the following questions.
Your earnings will not depend on the answers given to these questions.
***

The questions themselves are designed not to induce cognitive fatigue. For instance:

***

How much time per day do you spend reading the news?
Which topics do you spend more time reading about?
Which of these is the closest to your favorite color?

There are two different shapes in this image. Which ones are they? (*The subjects had the choice of leaving this question blank.*)

***

The second phase of the experiment consists of the actual games, in the same order in which the plans of action were elicited. They play the first (simpler) game, which they will see until they choose to quit (by choosing 39), or until they win, or until they lose all their tokens. Before playing either game, we inform them again that their earnings will be determined at random. In particular, the instructions explain that at the end of the experiment a die will be tossed to determine the earnings. If it lands between 1 and 3 then one game will determine their earnings, otherwise the other will.

After they are done with the first game, they can play the roulette game. We provide them with a sheet with the rules of the roulette (same as explained above), and they are free to raise their hands if they have any doubts, which almost never occurred. The subjects could quit at any moment, and they could also bet 0 tokens if they wished. The game can be found at: http://experimentalgames.upf.edu/roulette/ ; we include snapshots above in Figure 1. The roulette was coded in Adobe Flash.

The third phase of the experiment consists of the questionnaire. We ask them the 8 short grit questions, followed by the shortened 17 locus of control questions. We also ask them self-esteem, temptation and procrastination questions:

***

I have high self-esteem.

- 1 - Completely disagree.
- 2 - Somewhat disagree.
- 3 - Somewhat agree.
- 4 - Very much agree.
- 5 - Completely agree.

There may be tasks that you have to perform but that are not fun to do. For example, this could include studying courses that you dislike, waking up early, etc. Do you find yourself postponing these tasks or performing them less often than you should?

- 1 - Never
- 2 - Almost never.
- 3 - Sometimes.
- 4 - Often.
- 5 - A lot.

Similarly, there may be activities that you should not do too often, but which you enjoy doing. For example, this could include spending the day watching episodes of TV series, eating excessive amounts of chocolate, etc. Does it happen to you often that you prioritize or end up spending too much time doing these activities? Do you find yourself putting these activities ahead of more important ones or spending too much time performing them?

- 1 - Never
- 2 - Almost never.
- 3 - Sometimes.

- 4 - Often.
- 5 - A lot.

***

The last questions ask for their age, gender and field of study. Once a subject finishes the survey, he or she enters the control room (individually, to avoid social effects) and can choose to toss the die to determine which game will matter for the earnings.

B Further Robustness Estimations

Table 13: Robustness to Baseline Regression

Note: An Ordinary Least Squares specification is used. The dependent variable is a dummy variable taking the value 1 in case the subject overplayed or could potentially have overplayed (see Footnote 14). Grit Index takes values between 1 and 5 and increases with the level of grit. Plan of Action is the log of 1 plus the planned minimum bound. Standard errors are robust to heteroskedasticity. Significance: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Dep. Var.: Potential Overplaying	(1)	(2)	(3)	(4)	(5)
Grit Index	0.209***	0.197***	0.185**	0.202***	0.212***
	(0.074)	(0.072)	(0.073)	(0.071)	(0.072)
Plan of Action		0.070***	0.072***	0.078***	0.080***
		(0.012)	(0.012)	(0.014)	(0.014)
ln(Age)			0.278	0.143	0.072
			(0.269)	(0.256)	(0.254)
D(Female)				-0.202**	-0.231***
				(0.083)	(0.084)
Technical Degree					-0.086
					(0.083)
Constant	-0.301	-0.696***	-1.519*	-1.083	-0.852
	(0.252)	(0.248)	(0.836)	(0.802)	(0.794)
Observations	138	138	138	138	138
R-squared	0.053	0.114	0.121	0.160	0.166

Table 14: Different levels of final gains

Note: This table reports regressions similar to those in column (5) of Table 3. Column (1) excludes individuals with final gains above 6000 tokens. Column (2) additionally excludes individuals with final gains between 2000 and 6000 tokens. Finally, column (3) only keeps individuals with final gains below 1000 tokens. See also notes to Table 3. Significance: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

	(1)	(2)	(3)
Dep. Var.: Overplaying	<6000	<2000	<1000
Grit Index	0.264***	0.299***	0.239**
	(0.075)	(0.098)	(0.094)
Plan of Action	0.074***	0.079***	0.140***
	(0.014)	(0.016)	(0.020)
ln(Age)	0.045	-0.091	0.428
	(0.266)	(0.376)	(0.394)
D(Female)	-0.161*	-0.160	-0.133
	(0.083)	(0.100)	(0.106)
Technical Degree	-0.081	-0.011	0.094
	(0.082)	(0.100)	(0.123)
Observations	133	93	40
R-squared	0.173	0.199	0.525

Table 15: Different plans of action

Note: This table reports regressions similar to those in column (5) of Table 3. Column (1) excludes individuals with a plan of action of 2000 tokens. Column (2) additionally excludes individuals with plans of action between 1500 and 2000 tokens. Finally, column (3) only keeps individuals with plans of action below 1000 tokens. See also notes to Table 3. Significance: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Dep. Var.:	(1)	(2)	(3)
Overplaying	<2000	<1500	<1000
Grit Index	0.240***	0.233***	0.264***
	(0.074)	(0.081)	(0.086)
Plan of Action	0.057***	0.049***	0.039**
	(0.014)	(0.014)	(0.015)
ln(Age)	0.063	0.119	-0.145
	(0.274)	(0.284)	(0.306)
D(Female)	-0.166*	-0.148	-0.117
	(0.084)	(0.092)	(0.108)
Technical Degree	-0.062	-0.010	-0.046
	(0.082)	(0.085)	(0.100)
Observations	129	109	70
R-squared	0.141	0.132	0.155

Table 16: Adding temptation and procrastination as control variables

Note: This table resembles the baseline Table 3, but now also adds temptation and procrastination as control variables. Significance: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. See also notes to Table 3.

	(1)	(2)	(3)	(4)
Dep. Var.: Overplaying				
Grit Index		0.308***	0.317***	0.348***
		(0.087)	(0.084)	(0.079)
Procrastination	-0.046	0.029	0.048	0.038
	(0.062)	(0.061)	(0.059)	(0.060)
Temptation	-0.013	0.030	0.043	0.063
	(0.060)	(0.055)	(0.054)	(0.057)
Plan of Action			0.071***	0.084***
			(0.016)	(0.015)
ln(Age)				0.056
				(0.269)
D(Female)				-0.257***
				(0.084)
Technical Degree				-0.146*
				(0.084)
Observations	118	118	118	118
R-squared	0.008	0.099	0.158	0.226

Table 17: Overplaying >100 tokens

Note: This table uses some of the specifications of Table 3, but drops the 10 subjects that overplayed by less than 100 tokens. Significance: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

	(1)	(2)	(3)
Dep. Var.: Overplaying			
Grit Index	0.200***	0.187***	0.203***
	(0.074)	(0.071)	(0.073)
Plan of Action		0.070***	0.078***
		(0.013)	(0.014)
ln(Age)			0.110
			(0.257)
D(Female)			-0.184**
			(0.083)
Technical Degree			-0.114
			(0.081)
Observations	128	128	128
R-squared	0.057	0.129	0.171

Table 18: Alternative econometric specifications

Note: This tables uses alternative econometric specifications. Columns (1) and (2) uses a Logit specification, where the latter column is otherwise equivalent to column (5) of Table 3. Columns (3) and (4) proceed similarly with a Probit estimator. Finally, columns (5) and (6) use a Poisson specification. Significance: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Dep.Var.: Overplaying	(1) Logit	(2) Logit	(3) Probit	(4) Probit	(5) Poisson	(6) Poisson
Grit Index	1.281***	1.322***	0.751***	0.785***	0.823***	0.764***
	(0.412)	(0.454)	(0.239)	(0.256)	(0.252)	(0.252)
Plan of Action		0.631***		0.360***		0.405***
		(0.244)		(0.126)		(0.137)
ln(Age)		0.506		0.293		0.311
		(1.327)		(0.812)		(0.530)
D(Female)		-0.768*		-0.467*		-0.385*
		(0.428)		(0.259)		(0.221)
Technical Degree		-0.560		-0.311		-0.324
		(0.428)		(0.255)		(0.228)
Observations	138	138	138	138	138	138

Table 19: Robustness to alternative econometric specifications

Note: This tables uses alternative econometric specifications. The dependent variable is a dummy variable taking the value 1 in case the subject overplayed or could potentially have overplayed (see Footnote 14). Columns (1) and (2) uses a Logit specification, where the latter column is otherwise equivalent to column (5) of Table 3. Columns (3) and (4) proceed similarly with a Probit estimator. Finally, columns (5) and (6) use a Poisson specification. Significance: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Dep.Var.: Potential Overplaying	(1) Logit	(2) Logit	(3) Probit	(4) Probit	(5) Poisson	(6) Poisson
Grit Index	0.920**	0.997**	0.559***	0.611***	0.548***	0.504**
	(0.359)	(0.394)	(0.215)	(0.230)	(0.211)	(0.202)
Plan of Action		0.537***		0.321***		0.330***
		(0.165)		(0.088)		(0.096)
ln(Age)		0.451		0.272		0.237
		(1.244)		(0.775)		(0.460)
D(Female)		-1.089***		-0.672***		-0.521***
		(0.420)		(0.254)		(0.200)
Technical Degree		-0.458		-0.270		-0.248
		(0.409)		(0.246)		(0.204)
Observations	138	138	138	138	138	138

Table 20: Adding Locus Index as a control variable

Note: This table resembles the baseline Table 3, but now also adds the Locus Index. This index is defined in the range between 0 and 1, where higher values mean that the subject believes to a larger extent that outcomes in life are driven by external factors. Significance: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Dep. Var.: Overplaying	(1)	(2)	(3)	(4)
Grit Index		0.246***	0.236***	0.254***
		(0.073)	(0.071)	(0.072)
Locus Index	-0.373	-0.227	-0.196	-0.125
	(0.226)	(0.229)	(0.223)	(0.221)
Plan of Action			0.066***	0.073***
			(0.013)	(0.014)
ln(Age)				0.047
				(0.261)
D(Female)				-0.150*
				(0.083)
Technical Degree				-0.098
				(0.081)
Observations	138	138	138	138
R-squared	0.018	0.093	0.149	0.175

Table 21: Robustness to Larger Sample

Note: An Ordinary Least Squares specification is used. The dependent variable is a dummy variable taking the value 1 in case the subject overplayed. Grit Index takes values between 1 and 5 and increases with the level of grit. Plan of Action is the log of 1 plus the planned minimum bound. The last equation only includes data from the 3 sessions in which the Grit Questionnaire was introduced before the actual roulette playing. Standard errors are robust to heteroskedasticity. Significance: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Dep. Var. Overplaying	(1)	(2)	(3)	(4)
Grit Index	0.183***	0.182***		
	(0.065)	(0.064)		
Tenacity Index			0.253***	0.219*
			(0.063)	(0.126)
Diligence Index			-0.060	-0.178*
			(0.058)	(0.091)
Plan of Action		0.084***	0.082***	0.113***
		(0.013)	(0.012)	(0.038)
ln(Age)		0.121	0.174	0.680
		(0.240)	(0.256)	(0.556)
D(Female)		-0.118	-0.111	-0.026
		(0.072)	(0.071)	(0.141)
Technical Degree		0.001	0.033	0.299**
		(0.072)	(0.073)	(0.137)
Constant	-0.266	-1.084	-1.226	-2.575
	(0.217)	(0.745)	(0.802)	(1.738)
Observations	186	186	186	48
R-squared	0.045	0.148	0.176	0.287

Table 22: Appendix - Confirmatory Factor Analysis (CFA) - split of the grit index

Note: F1 includes the questions assigned to the Tenacity Index; F2 includes the questions assigned to the Diligence Index.

Item Content	F1	F2	Mean	Std. Dev.
Setbacks don't discourage me	0.099		2.898	0.961
I finish whatever I begin	0.578		3.275	1.016
New ideas and projects sometimes distract me from previous ones	0.592		2.956	0.988
I have been obsessed with a certain idea or project for a short time but later lost interest	0.653		3.203	1.033
I often set a goal but later choose to pursue a different one	0.603		3.746	0.793
I am a hard worker		0.362	3.696	0.859
I am diligent		0.527	3.609	0.899
I have difficulty maintaining my focus on projects that take more than a few months to complete		0.755	3.645	0.911

Table 23: Robustness to splitting Grit Index into tenacity and diligence (alternative split)

Note: This tablesplits Grit Index into the reduced measure of Tenacity Index and Diligence Index, but now only using 2 questions for each index. Column (1) uses the full sample of observations with no controls, similar to column (1) on Table 3. Column (2) adds the full battery of control variables, as in column (5) on Table 3. Columns (3) to (6) split the data by educational degree and gender. Finally, column (7) drops the 10 subjects who overplayed by less than 100 tokens. Significance: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Dep. Var.: Overplaying	(1)	(2)	(3) Technical	(4) Non-technical	(5) Female	(6) Male	(7) >100
Tenacity Index	0.159***	0.175***	0.216***	0.147*	0.210**	0.144*	0.140**
	(0.059)	(0.056)	(0.070)	(0.087)	(0.084)	(0.086)	(0.059)
Diligence Index	0.015	0.026	0.018	0.001	0.050	-0.019	0.015
	(0.061)	(0.059)	(0.085)	(0.091)	(0.077)	(0.093)	(0.062)
Plan of Action		0.084***	0.078***	0.098***	0.073**	0.079***	0.086***
		(0.014)	(0.019)	(0.026)	(0.029)	(0.017)	(0.015)
ln(Age)		0.122	-0.578	0.691***	-0.350	0.623**	0.162
		(0.287)	(0.436)	(0.258)	(0.400)	(0.266)	(0.281)
D(Female)		-0.143*					-0.177**
		(0.084)					(0.086)
Technical Degree		-0.107					-0.123
		(0.085)					(0.087)
Constant	-0.198	-1.045	0.855	-2.818***	0.078	-2.381**	-1.046
	(0.227)	(0.888)	(1.347)	(0.803)	(1.167)	(0.898)	(0.874)
Observations	138	138	71	67	80	58	128
R-squared	0.068	0.173	0.176	0.197	0.167	0.199	0.168

Table 24: ICPSR - Effect of Grit on Education

Note: The three dependent variables correspond to test scores obtained in the ACT Quality Core, the SAT9 exam, and the maths exams between 2009 and 2011. Columns (1)-(3) have the ACT Quality Core test score as the dependent variable. In the first column, the main variable of interest is the Grit Index. In the next two column, the index is decomposed into the Diligence Index and the Tenacity Index, the former including district fixed effects and the later rather including school fixed effects. Columns (4)-(6) replicate the same structure for the SAT9 exam, and columns (7)-(9) do it for the maths exam grade. Significance: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Dep. Var.:	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Test Scores	ACT	ACT	ACT	SAT9	SAT9	SAT9	Maths	Maths	Maths
Grit Index	1.641***			7.993***			0.166***		
	(0.175)			(0.774)			(0.009)		
Diligence Index		1.079***	1.058***		3.746***	4.133***		0.066***	0.068***
		(0.167)	(0.157)		(0.718)	(0.681)		(0.008)	(0.008)
Tenacity Index		0.540***	0.403**		4.244***	3.781***		0.099***	0.091***
		(0.193)	(0.176)		(0.795)	(0.764)		(0.009)	(0.009)
District FE	Y	Y	N	Y	Y	N	Y	Y	N
School FE	N	N	Y	N	N	Y	N	N	Y
Gender Dummies	Y	Y	Y	Y	Y	Y	Y	Y	Y
Age Control	Y	Y	Y	Y	Y	Y	Y	Y	Y
Race Dummies	Y	Y	Y	Y	Y	Y	Y	Y	Y
Additional Controls	Y	Y	Y	Y	Y	Y	Y	Y	Y
Observations	4,551	4,551	4,551	4,625	4,625	4,625	14,465	14,465	14,465
R-squared	0.228	0.230	0.339	0.319	0.322	0.413	0.333	0.333	0.407

Table 25: Online Survey - Effect of Grit on Conscientiousness

Note: In this table the dependent variable is the level of conscientiousness, ranging from 1 (lowest) to 5 (highest). Column (1) includes subjects from all over the world in a specification that only has the Grit Index and column (2) repeats the same specification with the Diligence and Tenacity Indices. Column (3) adds country fixed effects, and column (4) also controls for gender, race, and whether the subjects lives in an urban area. Column (5) and (6) limit the data to US respondents, without and with control variables, respectively. Significance: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Dep. Var.: Conscientiousness	(1)	(2)	(3)	(4)	(5) US only	(6) US only
Grit Index	0.617***					
	(0.012)					
Diligence Index		0.382***	0.388***	0.377***	0.404***	0.386***
		(0.014)	(0.014)	(0.015)	(0.021)	(0.021)
Tenacity Index		0.235***	0.231***	0.228***	0.221***	0.222***
		(0.015)	(0.016)	(0.016)	(0.022)	(0.022)
ln(Age)				0.126***		0.119***
				(0.025)		(0.032)
Gender Dummies	N	N	N	Y	N	Y
Race Dummies	N	N	N	Y	N	Y
Urban Dummies	N	N	N	Y	N	Y
Country FE	N	N	Y	Y	N	N
Observations	3,988	3,988	3,951	3,951	2,014	2,014
R-squared	0.410	0.430	0.453	0.458	0.417	0.424

Table 26: Online Survey - Effect of Conscientiousness and Grit on Education

Note: In this table the dependent variable is the level of education attained, ranging from 1 (lowest) to 5 (highest). Column (1) includes subjects from all over the world in a specification that only has the Conscientiousness Index. Column (2) only includes the Grit Index, while Column (3) includes both items. Column (4) splits the Grit Index into the Diligence and Tenacity Indices, and Column (5) additionally incorporates the Conscientiousness Index. Columns (6)-(10) replicate the same regressions after having added controls for age, gender, race, and whether the subjects live in an urban area, in addition to adding country fixed effects. Significance: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Dep.Var.: Education	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Conscientiousness Index	0.227***		0.114***		0.102***	0.089***		0.039*		0.027
	(0.019)		(0.025)		(0.025)	(0.016)		(0.021)		(0.021)
Grit Index		0.240***	0.170***				0.101***	0.078***		
		(0.018)	(0.024)				(0.015)	(0.019)		
Diligence Index				0.158***	0.118***				0.100***	0.089***
				(0.021)	(0.023)				(0.017)	(0.019)
Tenacity Index				0.083***	0.059**				0.004	-0.002
				(0.023)	(0.024)				(0.019)	(0.019)
Gender, Race, Urban FE	N	N	N	N	N	Y	Y	Y	Y	Y
Country FE	N	N	N	N	N	Y	Y	Y	Y	Y
Observations	3,988	3,988	3,988	3,988	3,988	3,951	3,951	3,951	3,951	3,951
R-squared	0.035	0.042	0.047	0.044	0.048	0.435	0.437	0.437	0.439	0.439

C External validity: grit components and education

Below is the description of the educational datasets.

C.1 ICPSR dataset on educational outcomes

This subsection uses large-scale data from the Measures of Effective Teaching (MET) Project (Bill and Melinda Gates Foundation 2014) to assess how our new categorizations of grit into tenacity and diligence relate to educational outcomes. The MET study provides information between 2009 and 2011 across six districts: Charlotte-Mecklenburg (North Carolina) Schools, Dallas (Texas) Independent School District, Denver (Colorado) Public Schools, Hillsborough County (Florida) Public Schools, Memphis (Tennessee) City Schools, and the New York City (New York) Department of Education.

This dataset is the result of an expansive effort to capture and analyze a variety of measures that could be used to evaluate teachers and provide them with feedback about their professional practice. Standard administrative and achievement data was collected, and students also completed a questionnaire that includes precisely the same eight questions on grit used in our experiment.

The scale of the MET Project affords a unique opportunity to examine variation across grades, subgroups and classrooms. Although the data collected by the MET Project are not a random sample of the national population, they are broadly representative of the population of urban public school students across much of the United States. For our sample of students in Grades 6 to 9, we focus on scores on the standardized ACT test (N=5,733), the standardized SAT9 test (N=5,875), and a standardized Math test (N=18,088) as outcome variables. Together, these tests span a broad range of topics, from more mathematical and technical ones to more reading-oriented ones.

Table 24 summarizes the results on these three dependent variables. Column (1) presents results with the ACT Quality Core test score as the dependent variable and the Grit Index as the main variable of interest. The estimated coefficient of 1.641 is statistically significant and in economic terms it means that a one unit increase in the Grit Index is associated with an increase in the dependent variable of one fifth of a standard deviation. In the next column we decompose the index into its Diligence and Tenacity components. Both indices are statistically significant and positively correlate with the ACT test score. Similar statistical and economic results are found when we use instead the SAT9 test score as the dependent variable (columns (4)-(6)) or the average of a student's math scores in the period 2009-2011 (columns (7)-(9)). These results show that both tenacity and diligence are indicative of higher educational performance in various domains.

C.2 Cross-country online survey of educational attainment

This dataset consists of a sample of 3988 individuals which includes data on grit, conscientiousness, and education as well as other demographics.[23] Respondents are informed before taking the test that the data may be used for research purposes. The survey can be taken online by anyone who chooses to do so and it is anonymous. While all the answers are self-reported, including educational outcomes, we do not expect any bias in our estimation. This dataset is particularly informative because the respondents have a wide heterogeneity in educational level, age, ethnic background and geography.

The mean of the Grit Index for this dataset is 3.25 (3.31 for the US), which is in line with the mean in our experiment (3.38). This similarity also holds when splitting the questions into

[23] This data is drawn from the online psychology survey repository available at personality-testing.info/_rawdata (March 17, 2016). This archive of psychological tests has been used in several articles in the psychology literature.

tenacity and diligence: the mean for diligence (tenacity) is 3.62 (3) in this dataset compared to 3.65 (3.22) in our experiment.

While we focus on grit because it incorporates the tenacity category which is central to our hypothesis, grit and conscientiousness have been found to be highly related, which raises the point of whether our results hold with conscientiousness as well. For this reason, in Table 25 we use the online survey to study the connection between the different categories of grit and conscientiousness. Using conscientiousness as the dependent variable, we first confirm it is highly correlated with grit, in both an economic and statistical sense. When decomposing grit into diligence and tenacity, we again find that both are highly positively correlated, but that tenacity does not entirely capture conscientiousness. This suggests that while we do expect that our main findings on the flip side of tenacity may be present in conscientiousness as well, the two are sufficiently distant that understanding more precisely the degree to which our results extend to conscientiousness requires further exploration.

We now document how educational outcomes are explained by the Conscientiousness Index and Grit Index, and the split into Diligence and Tenacity Indices (Table 26). Columns (1)-(5) run the basic specifications, while columns (6)-(10) include country fixed effects and control for age, gender, racial background, and urbanization. In general, the Conscientiousness and Grit Indices are positive and statistically significant on their own and together (columns (1)-(3) and (6)-(8)). When regressing education on conscientiousness and diligence and tenacity, all three are significant without controls (column (5)), and diligence remains significant with controls as well. Conscientiousness is close to being significant, while the coefficient for tenacity effectively drops to zero (column (10)). The same pattern remains for diligence and tenacity without conscientiousness: they are both significant without controls, and only diligence is significant with controls.

These results confirm the upside of diligence discussed in our analysis, and shows that tenacity is more nuanced. The interplay between the upside and the downside of tenacity is likely to be intricate in a setting as complex and multidimensional as educational attainment. But disentangling the two sides would be of interest in future research. Our results are also in line with the findings in the literature that conscientiousness is predictive of educational performance, and at the same time show that our new categories are independently informative of performance as well.

D Confirmatory Factor Analysis explanation

Big Picture Summary of CFA and its link to our setup

Confirmatory factor analysis (CFA) is a multivariate statistical procedure used to test whether measures of a construct are consistent with a researcher's understanding of the nature of that construct (or factor). The measurable variables are called manifest variables. The unmeasurable are called factors (or latent variables). In our example, 'tenacity' would be the factor, while the subset of observable questions we believe represent this factor are the manifest variables.

As such, CFA is conducted to test theories and hypotheses about the factors one expects to find in the data. This hypothesized model is based on theory and/or previous analytic research. Consequently, CFA is a tool that is used to confirm or reject the measurement theory.

The researcher first develops a hypothesis about which factors they believe are underlying the measures used and may impose constraints on the model based on these a priori hypotheses. Researchers can specify the number of factors required in the data, and which measured variable is related to which latent variable. Therefore, CFA allows us to answer the following question: does my subset of questions of the Grit Index accurately measure one factor, in this case, 'tenacity'?

Assessing the Model Validity

Assessing the measurement model validity occurs when the theoretical measurement model is compared with the reality model to see how well the data fits. If the constraints the researcher has imposed on the model are inconsistent with the sample data, then the results of statistical tests of model fit will indicate a poor fit, and the model will be rejected.

With regard to selecting model fit statistics to report, several statistics are available including the comparative fit index (CFI), the Tucker-Lewis index (TLI), or the standardized root mean squared residual (SRMR).