Are Humans More Averse to Aggregate Risk? Testing an Evolutionary Economic Theory *

Ryan Oprea † Nikolaus Robalino‡

August 15, 2024

Abstract

We experimentally test a classic prediction of evolutionary economics: that natural selection should have favored decision makers with greater aversion to aggregate risk (risks correlated with those of others) than to idiosyncratic risk. In a series of risk elicitation tasks, we vary the correlation between subjects' risks with risks faced by 30 other participants in several distinct ways. In line with the theory, we find that inducing positive cross-subject payoff correlations consistently causes subjects to make choices that are significantly more risk averse than in baseline tasks without correlated payments, and that a placebo task inducing negative correlation has no such effect. We also find strong evidence of aggregate correlation aversion, which is predicted by the evolutionary theory: subjects prefer to avoid decision settings in which their payoffs are correlated with those of others. Collectively, behavior across our treatments forms

*This paper is dedicated to the memory of John Tooby, a pioneer in the application of evolutionary reasoning to the understanding of human behavior. We thank Theodore Bergstom, Daniel Friedman, Yuval Heller, Ilan Nehama, Arthur Robson, Philipp Sadowski, Larry Samuelson, and Jakub Steiner for very helpful comments. This research was supported by the National Science Foundation under Grant SES-1949366 and was approved by UC Santa Barbara IRB.

†Oprea (corresponding author): Economics Department, University of California, Santa Barbara, Santa Barbara, CA, 93106, roprea@gmail.com.

‡Robalino: Economics Department, Rochester Institute of Technology, Rochester, NY, 14623, ndrobalin@gmail.com.

a distinctive empirical fingerprint that matches this theory, but not alternative explanations (e.g., social learning or social preferences).

Keywords: evolution, natural selection, risk preferences, aggregate risk, idiosyncratic risk, correlation aversion

JEL codes: C91, D91 G0, K4

1 Introduction

In standard economic theory, decision makers' risk preferences are insensitive to correlations between their payoffs and the payoffs of others. An investor's personal tolerance for risk, for example, should not change upon learning that others' payoffs are positively correlated with her own. However, an important line of theoretical research studying how evolutionary forces might have shaped human risk preferences suggests otherwise. A long economics literature, beginning with Robson (1996), argues that agents with greater aversion to aggregate (positively correlated) risk than to idiosyncratic (uncorrelated) risk are more reproductively fit, and therefore will have been favored by natural selection. The evolutionary theory therefore suggests that human risk preferences should differ fundamentally in aggregate settings featuring shared risks (e.g., investment in the stock market, voting behavior) than in, e.g., individually isolated decision-making featuring only idiosyncratic risk (e.g., preference elicitations, health decisions, consumer choice).

This evolutionarily-derived hypothesis—call it the "aggregate risk hypothesis" or ARH—if correct, has important implications for economics. Perhaps most importantly, it suggests severe limits on the use of revealed preferences for risk in applied work: the same agent may reveal significantly more risk averse preferences in their voting or investing behavior (choices beset by aggregate risk) than they would make in consumer purchases, private health decisions or in direct preference elicitation (choices involving mostly idiosyncratic risks). As a result, revealed risk preferences based on idiosyncratic risky choice will do a poor job of calibrating or predicting behavior in aggregate settings like macro-economics or political economy, since they are based on flawed counterfactuals. Indeed, the ARH offers a resolution to the "equity premium puzzle" (Robson and Orr, 2021), the famous finding that people invest less in equity than measures of idiosyncratic risk aversion suggest they will (Mehra and Prescott, 1985). The ARH might also explain phenomena such as the reputed sluggishness and conservatism of large organizations, polities and committees relative to smaller upstarts (larger groups may be less tolerant of risk because, by virtue of their size, they face more aggregate risk).

Despite its potential importance, the ARH has yet to be included in the corpus of models typically used by behavioral economists to explain economic behavior. This is doubtless because of perennial concerns about the relevance of evolutionary hypotheses to contemporary economic behavior. After all, evolutionary theories such as the ARH are derived from partial equilibrium models that capture only a small number of the many evolutionary forces that might have shaped contemporary human behavior. The theory alone cannot verify that the selective pressures in favor of the ARH survived competing selective forces in general equilibrium. Moreover, the ARH is, strictly speaking, a population-level hypothesis describing how humans will respond to risks that are shared by the entire human population. It is not clear that the implications of the ARH should apply also to the smaller-scale groups, such as nations, cities, firms, or experimental cohorts that are the focus of most applied work. Finally, although the theory suggests that agents should have an aversion to aggregate risk, the behavioral mechanism by which this aversion was implemented needn't have been risk preferences, per se. For instance, avoidance of aggregate risk may instead have been produced by the evolution of status concerns or other forms of social preferences. These ambiguities are difficult to resolve using theoretical reflection alone, and can only be overcome by subjecting the ARH to a direct empirical test.

In this paper we report a demanding experimental test of the ARH. Our motivation for this test is precisely the fact that it *isn't* clear ex ante whether the ARH should show up in evolutionary novel settings like an experiment, making an experimental test particularly valuable for understanding the scope and strength of the ARH. The experiment is built on novel variations of the "bomb risk elicitation task" (BRET), an intuitive graphical method for eliciting risk preference parameters (Crosetto and Filippin, 2013). In our baseline ID (i.e., "idiosyncratic risk") treatment subjects choose a number, $b \in \{1, 2, ...50\}$ after which a state $s \in \{1, 2, ...50\}$ is randomly and uniformly selected by a computer. The subject earns b if $b < s$, and earns 0 otherwise. A choice of $b = 25$ reveals risk neutrality but risk averse subjects should choose lower values of b; a CRRA parameter can be calculated directly from this choice. In our experiment, conducted on a demographically diverse, primarily non-student online subject pool, we find standard measured levels of risk aversion, with an average implied CRRA parameter of 0.738 in the ID task (given the CRRA utility, $v(x) = x^{\mu}$).

In a series of additional treatments, run on the same subjects, we introduce aggregate risk to this task in several distinct ways. In the POS (positive correlation) treatment, subjects observe the b choice of 30 other subjects prior to making their choice and are truthfully told that the same state s will be selected for all subjects, producing positive correlations in earnings across subjects, including the decision maker. In the LEAD (leader) treatment, the subjects is instead told that her choice of b will directly determine the b for 30 other subjects; again, all subjects receive the same draw of s, this time producing perfect positive correlation in payoffs. In both of these treatments, the ARH predicts a significant increase in risk aversion (e.g., a lowering of the CRRA parameter) relative to the baseline ID treatment.

We find strong evidence in favor of the ARH in both of these treatments. Relative to the ID treatment, we find that subjects in both the POS and LEAD treatments make significantly *more* risk averse choices. Our test is particularly stringent because we use a within-subjects design: subjects in the POS and LEAD treatments also make decisions in the baseline ID treatment (in a random order) and so we can study whether individual subjects change their decisions across treatments. We find that the mean subject adjusts her LEAD and POS choices in a significantly more risk averse direction relative to her ID choice. For the average subject the CRRA parameter is 0.12 lower in the LEAD and POS treatments than in the ID treatment—a sizable and statistically significant increase in aversion to risk (equivalent in magnitude to estimates of the gender gap in risk preferences, e.g., Holt and Laury (2002)).

The interpretation of these results is aided by a placebo treatment that tests a distinctive negative implication of the ARH: that these increases in risk aversion will disappear in settings where the decision maker's payoff is *negatively* rather than positively correlated with the payoffs of the 30 other subjects. In the NEG (negative correlation) treatment, we mirror the POS treatment except the decision maker draws a state s that is instead perfectly *negatively* correlated with the state drawn for the other subjects.[1] This treatment is cosmetically nearly identical to the POS treatment. However, the ARH implies there should be no increase in risk aversion in NEG relative to ID and, in fact, predicts a reversal of the result from POS. This is just what we find. In sharp contrast to the POS treatment, we find a slight *decrease* in risk aversion in the NEG treatment relative to the ID treatment.

In a final empirical step, we directly test the ARH's prediction that individuals will be averse to positively correlated payoffs across subjects. At the end of the experiment, we ask subjects to choose which treatment to make a final risky choice in: ID vs. LEAD/POS/NEG. We find that when given the chance to choose an environment, subjects are more than twice as likely to choose the environment that induces less correlation in payoffs. What's more, we find that the increased risk aversion displayed in POS and LEAD only occurs for these correlation averse subjects, further supporting the mechanism described by the ARH.

The ARH predicts a distinctive fingerprint across the LEAD, POS and NEG treatments, meaning our design jointly poses a stringent test of the ARH relative to alternative explanations for our results. For instance, while some forms of social learning might explain the change of behavior we observe in POS, they cannot explain the change in behavior we observe in LEAD or the null effect we observe in NEG. Likewise, while some types of social preferences might explain the decrease in risk aversion observed in LEAD, they cannot explain the symmetric decrease observed in POS (where subjects choices do not impact others' payoffs). Conversely, while status concerns or preferences over relative consumption might explain the apparent increase in risk aversion that we find in POS, these types of social preferences cannot alone explain the identical increase in risk aversion that we observe in LEAD. Our design therefore poses a demanding test of the ARH and we find that the ARH parsimoniously predicts both the increases in risk aversion that we observe in LEAD and

[1] The other subjects continue to draw a common state, hence their payoffs remain positively correlated as in POS, though they are negatively correlated with the payoff of the decision making subject.

POS, the absence of such an effect in NEG, and the connection between these results and the aversion to correlated payoffs we find in the data.

These results reveal an important characteristic of risk preferences with significant implications and applications, especially for an aggregate science such as economics. However, in our concluding discussion, we also argue that our findings suggest a promising methodological program for further understanding economic behavior—the use of a growing body of evolutionary models in economics as a source of testable *behavioral hypotheses*. Traditionally economists have developed testable behavioral hypotheses via appeal to normative axioms, casual empiricism or through the adaptation of findings from other fields like psychology. By testing and calibrating these hypotheses in the field or in the lab, we have refined them over time, improving over time our ability to do our job of predicting and explaining behavior and of designing effective policies and institutions. We argue that our results suggest that economics may benefit from adding evolutionary models to this production process.

Our paper is organized as follows. In Section 2, we discuss the evolutionary theory that underlies the ARH. In Section 3 we discuss the experimental design and in Section 4 describe a set of ten hypotheses implied by the ARH. In Section 5 we report our empirical results and in Section 6 we conclude.

1.1 Previous Literature

Our paper relates to three main economics literatures. First is a large literature in behavioral economics documenting departures from standard expected utility theory (EUT), often using experimental tests. These empirical findings have been used to produce a number of non-EUT models of preferences like prospect theory (Kahneman and Tversky, 1979; Tversky and Kahneman, 1992)). Relative to the bulk of this literature, our work is novel in that it takes an evolutionary model as the starting point and motivation for testing non-EUT risk preferences.

Second is a theoretical literature in economics (the "evolutionary foundations literature") that uses evolutionary models in order to understand the origins of economic preferences and behavior. The strategy adopted in this literature is to model simple evolutionary settings in which economically relevant behaviors have implications for reproductive fitness. By identifying which behaviors are reproductively adaptive, the literature forms hypotheses about the structure of economic primitives such as preferences; see Robson (2001a), Robson (2002), and Robson and Samuelson (2011) for surveys of this literature. This literature has used this method to study risk preferences (e.g., Robson, 1996), time preferences (e.g., Robson and Samuelson, 2009), preferences for bet-hedging (e.g., Cooper and Kaplan, 1982; Bergstrom, 1997, 2014), probability weighting (e.g., Steiner and Stewart, 2016), strategic reasoning and theory of mind (e.g., Mohlin, 2012; Robalino and Robson, 2016; Kimbrough, Robalino and Robson, 2017), hedonic utility, (e.g., Friedman, 1989; Robson, 2001a; Rayo and Becker, 2007; Netzer, 2009), and the origins of utility functions (e.g., Robson, 2001b).[2,3] To our knowledge, Kimbrough, Robalino and Robson (2017) is the only previous economics experiment that is explicitly motivated by the evolutionary foundations literature, though they study a very different phenomenon than we do (the evolution of theory of mind).

Among these, most related to our work is a literature on the evolutionary foundations of risk preferences. This literature began with Robson (1996) (reviewed in Section 2) who showed in a general model that natural selection will have favored agents with greater aversion to aggregate risks than to commensurate idiosyncratic risks (see Lewontin and Cohen (1969) for an earlier, related contribution). Curry (2001) considers a direct utility implementation of the ARH by modeling the evolution of utility functions over offspring when the environment entails both idiosyncratic and aggregate risk. The evolutionarily fittest individuals choose gambles that maximize expected relative numbers of offspring, given the choices of the others in the population. Such individuals will display behavior consistent with the evolutionarily optimal attitudes to aggregate and idiosyncratic risk. Alternatively, Heller and Nehama (2023) considers a "dual utility" implementation of the ARH. Individuals use one utility function to assess idiosyncratic gambles, and another, more risk averse, utility function

[2]This literature is, of course, related to large literatures in psychology (e.g., Barkow, Cosmides and Tooby, 1992) and anthropology (e.g., Boyd and Richerson, 1985) that have theoretically and empirically studied the relationship between evolution and human behavior.

[3]More distantly related to our paper is a large literature in economics on the use of evolutionary concepts to understand large scale behavior in population games (e.g., Friedman, 1991).

when risks are correlated. Utility is no longer interdependent in the evolutionary optimum, as it is in Curry (2001); instead there is heterogeneity in the risk attitudes corresponding specifically to correlated risks. We designed our experiment to largely abstract from these questions of implementation. Sadowski and Sarver (2023) show that adding hidden actions or phenotypic flexibility to a model like Robson (1996) can produce evolutionarily optimal behaviors that resemble ambiguity aversion and non-expected utility.

One strand of the evolutionary risk literature that is a particularly strong motivation for our study is a recent series of papers showing that the emergence of aggregate risk preferences depends on how we model evolutionary time. For example, Robatto and Szentes (2017) finds there is no difference in attitudes to aggregate and idiosyncratic risks when births arrive in continuous time, when there is no gap between birth and menarche. On the other hand, Robson and Samuelson (2019) shows that attitudes to aggregate and idiosyncratic risk will differ in continuous time models for populations with a richer age structure. Indeed, in another continuous time setting Robson and Samuelson (2022) shows that aggregate risk aversion can be arbitrarily high when fertility is attainable only once agents meet an exogenous consumption threshold. In another continuous time setting, Heller and Robson (2021) shows that skewness loving, as well as aversion to aggregate risk, is evolutionarily optimal. Interestingly, they also show that although individuals ought to be averse to correlation in payoffs within their own cohorts, they will prefer own risk to be correlated with the risks faced by their parents. The sensitivity of aggregate risk aversion to the details of the evolutionary environment is another strong motivation for conducting an empirical test like ours.

Finally, there is an experimental literature examining risk taking in social contexts. Much of this work is focused on how individual risky choices differ from risky choices by groups, or how risky choices for self differ from choices made on behalf of others. For example, an early observation is that risky choices by groups diverge from choices made by individual members for themselves, though results on the direction of the risk shift is mixed.[4,5] Relatedly, a literature examines differences in risky choices for self, and risky choices for others. The meta-analysis Polman and Wu (2020), which considers 71 articles, finds a small but significant difference suggesting that individuals make more risky choices for others than they do for themselves.[6]

There are few papers that examine cross-population correlations and risk preferences. Two that are, perhaps, closest to our work are Bolton, Ockenfels and Stauf (2015) and Rohde and Rohde (2011). Bolton, Ockenfels and Stauf (2015) ask subjects to make decisions on behalf of another subject under positive (much like our LEAD treatment) and negative correlations, in order to study the influence of inequality aversion on choice. They find that both treatments increase risk aversion relative to idiosyncratic risk—a result that provides only mixed support for the ARH, perhaps because Bolton, Ockenfels and Stauf (2015) study very small 2-person groups in which aggregate risk aversion plausibly shouldn't hold. In one of their treatments Rohde and Rohde (2011) find that decision makers prefer their lottery payoffs to be statistically independent of the payoffs of $N = 9$ other subjects, rather than positively correlated—a result related to our test of Hypothesis 6 that is consistent with the ARH. Neither of these papers are explicitly motivated by the ARH, nor are their experiments designed with a test of the ARH in mind. Relative to these papers, our experiment studies much larger cohorts in which the ARH more plausibly applies, and we study a much broader set of tasks that are designed specifically to pose a demanding test of the ARH.

[4]For example, Stoner (1961) finds that groups make more risky choices than individuals, while Nordhöy (1962) and Stoner (1968) find that individuals are more risk tolerant. The direction of shift appears to depend heavily on the context. For instance, Shupp and Williams (2008) find that groups are more risk averse than individuals in high-risk situations, but less risk averse for small risks.

[5]There is also a large literature on social preferences, which tests whether arguments of utility ought to include the payoffs, motives, or intentions of others. For example, see Kahneman, Knetsch and Thaler (1986); Loewenstein, Thompson and Bazerman (1989); Charness and Rabin (2002), and Andreoni and Miller (2002). For a survey see Cooper and Kagel (2015).

[6]In the experiments analyzed by Polman and Wu (2020) decision makers are observed making risky choices for themselves and also risky choices in a social context where their choice determines the risk faced by another individual. In only a small number of these ($k = 5$) does the choice in the social context determine risk for *both* the decision maker and the other individual, as in our LEAD treatment. In the remaining experiments, the social decision affects only the payoff of the other individual. In these cases it is not clear how (or if) attitudes to correlated payoffs might have played any role.

2 Theoretical Background

Robson (1996) considers a population of one-period-lived reproducing agents.[7] Individuals in each period face aggregate uncertainty in the form of a common environment, described by a state ω that fluctuates randomly from period-to-period. At each date, each agent chooses a reproductive gamble from a fixed set $\mathscr{G}$, where the distribution of outcomes produced by a gamble can depend on the state of the environment. In particular, suppose an agent has chosen the gamble $\pi \in \mathscr{G}$. In state ω this individual will have j surviving offspring with probability $\pi(j, \omega)$. Assume the number of surviving offspring is independent across individuals, conditional on the state of the environment.

An individual's choice of gamble from $\mathscr{G}$ is determined by her "type," inherited from her parent.[8] Individual choices of gambles determine the distribution of total surviving offspring of agents of a given type. In general this distribution will depend on the environmental state. Hence, there may be aggregate uncertainty regarding the average number of surviving offspring per individual, even when the number of parents is large. Suppose agents of type τ have J surviving offspring per capita with probability $P_{\tau n}(J \mid \omega)$, when there are n agents of type τ and the state of the environment is ω. Assume further that $P_{\tau n}$ has a well defined limit, P_τ, as n tends to infinity.[9] Suppose the state is governed by Markov transition probabilities with an ergodic distribution, Q. Robson (1996) shows that if the τ type avoids extinction (i.e. n does not reach 0), then the long-run growth rate of this type converges almost surely to

$$\Phi_\tau = \sum_\omega \log \left(\sum_J J \cdot P_\tau(J \mid \omega) \right) \cdot Q(\omega). \tag{1}$$

[7]Robson and Samuelson (2019) extend the results from Robson (1996) to more general age-structured populations.

[8]In Robson (1996) the type corresponds to a fixed choice of gamble. In Curry (2001) a type corresponds to how an individual's choice of gamble depends on the choices of others in the population, as in a game.

[9]If the type can avoid extinction in the short-run, then what matters is the type's behavior in the long-run. Stationarity in the long run follows, for example, when the type is associated with a fixed gamble (as in Robson (1996)), or a fixed utility function (as in Curry (2001)).

When the population is large $\sum_J J \cdot P_\tau(J \mid \omega)$ is offspring per individual given that the state is ω. Hence, the long-run growth rate of the τ type population is the expectation of the logarithm of offspring per individual of this type.

Consider types, τ and σ. If $\Phi_\tau > \Phi_\sigma$, then conditional on the τ-types surviving extinction in the long run, the σ-types will almost surely be overtaken by the τ-types. Hence, Robson (1996) argues that natural selection will favor the type with a limiting distribution over offspring per capita that maximizes (1).[10] In view of the eventual dominance of types that maximize (1), from now on, we refer to $\log(J)$ as Nature's utility function.

The first key observation made by Robson (1996) is that the concavity of log in (1) implies an aversion to risk in aggregate offspring as follows. Consider two types, τ and σ, such that the limiting distributions P_τ and P_σ have the same mean, say $\bar{J}$. Suppose, however, that for a τ-type individual the distribution of own offspring does not depend on ω, while for the σ-type the realized number of offspring does depend on the state. That is, the former type faces only idiosyncratic risk, while the latter type faces environmental (aggregate) risk. When the population is large, $\Phi_\tau = \log(\bar{J})$. Hence, the law of large numbers fully insures the τ population against idiosyncratic risk. On the other hand, the concavity of log gives

$$\begin{aligned}\Phi_\sigma &= \sum_\omega \log\left(\sum_J J \cdot P_\sigma(J \mid \omega)\right) \cdot Q(\omega) \\ &\leqslant \log\left(\sum_\omega \sum_J J \cdot P_\sigma(J \mid \omega)\ \cdot Q(\omega)\right) = \log(\bar{J}) = \Phi_\tau.\end{aligned} \tag{2}$$

Equality holds here only if $E_\sigma[J|\omega] = \bar{J}$, for each state of the environment, ω. Hence exposure to aggregate risk is evolutionarily disadvantageous in the sense that it implies a dominated growth rate. This remains true when P_τ and P_σ imply the same marginal distributions of offspring at the individual level. Thus, even when τ and σ agents incur risks that are indistinguishable from the individual's perspective, the type facing only idiosyncratic risk will have an evolutionarily edge, because in the long-run this

[10] The Robson (1996) model is general in that transmission of type from parent to offspring is noisy. When there is mutation the ultimate dominance of the type maximizing the above growth rate will follow in the limit of models where the probability of mutation tends to zero.

population will grow at a faster rate.

The second key observation of Robson (1996) is that the evolutionarily dominant type is less averse to idiosyncratic reproductive risk (risk uncorrelated with the risks of others) than to aggregate (correlated) risk. In particular, suppose individuals of type τ and σ choose gambles that yield statistically independent numbers of offspring across individuals. In this case, $\Phi_\tau > \Phi_\sigma$ if and only if P_τ has a higher mean than P_σ. Hence, in the absence of aggregate risk, the evolutionarily fittest type chooses the gamble that maximizes own expected offspring. In view of the aggregate risk aversion described above, the dominant type therefore displays a greater aversion to risk in aggregate offspring than to commensurate idiosyncratic risk.

These results concern attitudes to risk in offspring. However, as the literature has emphasized, if consumption is linked to fertility, then attitudes toward offspring risk will imply corresponding attitudes towards consumption risk. Thus, these findings will hold for risk more generally and will therefore have direct implications for economic behavior. We discuss this in further detail in the next section.

2.1 The Aggregate Risk Hypothesis

The theory sketched in the previous section suggests that evolution may have produced individuals with an aversion to population-wide correlated risks, with greater aversion to risk when it is aggregate (i.e. correlated across people) than when it is purely idiosyncratic. We will call this prediction the *aggregate risk hypothesis* (ARH), and evaluating it empirically is the purpose of our paper.

To operationalize the ARH and to produce crisp benchmark hypotheses for our experiment (discussed in Section 4), consider an evolutionary environment in which individuals choose gambles over consumption. Following Robson (1996), suppose consumption of $x \in [0, \infty)$ yields j offspring with a given exogenous probability, $p(j, x)$.[11] A choice of gamble now yields distributions over consumption, which might depend on the state of the environment. Assume that numbers of offspring are inde-

[11]The technology mapping consumption to offspring is independent of type, and independent of the size of the population.

pendent across individuals and independent of the environment, conditional on each realized profile of consumption levels.[12] Suppose, moreover, that more consumption unambiguously improves fertility:[13]

Assumption 1 *Given any levels of consumption x and $\bar{x}$, if $\bar{x} < x$, then the distribution $\{p(\cdot, x)\}$ first-order stochastically dominates $\{p(\cdot, \bar{x})\}$.*

To operationalize the ARH suppose, as in Curry (2001), a "type" of individual describes the way the individual conditions her own choice of gamble on the distribution of choices made by others. In the long-run, if the population avoids extinction, the evolutionarily dominant type is then one that chooses gambles over consumption that manages to ultimately maximize Nature's utility as given in (1). We now describe the attitudes to consumption that are implied by the ARH. Each of these results is proved in the Online Appendix.

First, under the ARH, decision makers will prefer less correlation in payoffs, ceteris paribus, producing a first testable implication of the theory:

Proposition 1 [**Aggregate Correlation Aversion (ACA)**] *The evolutionarily dominant type acts as if she has an aversion to correlation in hers and others consumption, and, moreover, as if she is averse to correlations in others consumption.*

The proof is given in Section **??** of the Online Appendix. ACA follows from the concavity of Nature's utility in aggregate offspring (aversion to risk in aggregate offspring), given Assumption 1. Affinity for correlation would follow if instead Nature's utility were convex in aggregate offspring.[14]

Next, the dominant type displays greater aversion to risk in *own consumption* when own consumption is more correlated with the consumption of others. This yields

[12] Since consumption may depend on the state, there may, however, be correlated consumption across individuals, and thus aggregate risk in offspring.

[13] We further impose some mild regularity assumptions on p (Assumption **??** in the Online Appendix).

[14] Attitudes to correlation in others' payoffs are revealed when, for example, an individual is observed making risky choices for others, or when the individual is allowed to choose the correlation structure of others' payoffs, as in our CHOICE treatments described below.

another testable implication of the theory (see Section **??** of the Online Appendix for the proof).

Proposition 2 [Aggregate Risk Aversion (ARA)] *The evolutionarily dominant type acts as if she has a greater aversion to risk in own consumption when own and others' consumption is more correlated.*

Both of these results use the following notions of "more correlated" and "correlation averse", which are due to Epstein and Tanny (1980). (Formal definitions are given in the Online Appendix, in Section **??**.) Let x_i denote the level of consumption of individual i in the population. Given individuals r and s in the population, consider joint distributions μ and η over (x_r, x_s). Say that η *induces more correlation* in x_r and x_s than does μ, if μ and η imply the same marginal distribution over x_i, $i = r, s$, but under η it is more likely to observe pairs (x_r, x_s) where x_r and x_s are both high, or both low.[15] An individual is then averse to correlation in r and s's consumption if, ceteris paribus, she prefers the distribution μ to η whenever η induces more correlation in x_r and x_s than does μ.[16]

These results suggest that attitudes to risk in own consumption will depend on the distribution of the payoffs of others, when own and others' payoffs are correlated. On the other hand, for the evolutionarily dominant type attitudes to *idiosyncratic* risk in own payoff are *independent* of the distribution of choices of others. This observation will play a role in generating sharp hypotheses for our experiment in Section 4:

Proposition 3 [Independence of Attitudes to Idiosyncratic Risks (IAIR)] *The evolutionarily dominant type acts as if her attitudes to idiosyncratic risk in own consumption are independent of the distribution of choices of others in the population.*

[15] That is, for each positive α, α',

$$\eta(\{x_r \geqslant \alpha, x_s \geqslant \alpha'\}) \geqslant \mu(\{x_r \geqslant \alpha, x_s \geqslant \alpha'\}),$$

and

$$\eta(\{x_r \leqslant \alpha, x_s \leqslant \alpha'\}) \geqslant \mu(\{x_r \leqslant \alpha, x_s \leqslant \alpha'\}).$$

[16] Correlation aversion persists in the large population limit. Fro example, when the population is large, individuals are strictly averse to own payoff being positively correlated with the payoffs of a non-negligible fraction of the population.

This is analogous to the result from Robson (1996) discussed above, that in the absence of aggregate risk the evolutionarily dominant type is a classical expected utility maximizer. IAIR is also evolutionarily optimal in the environments of Curry (2001), and Heller and Nehama (2023). The proof of the result is in Section **??** of the Online Appendix.

2.2 Challenges to the ARH

How relevant is the ARH to contemporary economic behavior? Our paper is motivated by the fact that this is not clear from the theory alone, for at least three reasons.

First, the models that motivate the ARH are, strictly speaking, *population* models that describe aggregate risks that afflict a large population. As such, it is not obvious that the ARH will apply to risks correlated within smaller economically relevant groups such as a family, a firm, a city, or even a country. Did nature install these preferences to be sensitive only to aggregate risks afflicting whole populations, or did it install more general preferences sensitive to aggregate risks faced by cohorts of *any* size?

Second, evolutionary models are unavoidably "partial equilibrium" models, focusing only on some small number of environmental factors (of many possible relevant ones) that determine reproductive fitness. In a fully specified model that accounts for the full suite of evolutionary forces shaping human behavior, tradeoffs, interactions and path dependencies between behavioral traits have the potential to intensify, attenuate or eliminate any manifestation of the ARH. Did the behavior implied by the ARH "survive" this clash of evolutionary forces?

Third, although Robson (1996) describes Nature's utility function, it does not directly specify how natural selection might have maximized this function. One possibility (the possibility the literature has focused on and that we test) is that it did so by shaping risk preferences to be correlation-sensitive—a possibility operationalized in, e.g., Curry (2001) and Heller and Nehama (2023).[17] Alternatively, however, ARH-

[17]Curry (2001), for example, proposes a direct implementation, in which (i) individual choices are guided by interdependent utility functions over own and others' offspring and (ii) each individual maximizes expected utility, given the choices of gambles of the others. The utility function of the

consistent choices might have been induced by evolution using a behavioral instrument other than risk preferences—for instance, some form of status concern, altruism, or perhaps by hardwired behaviors that are overfit to risks that happened to be correlated in the evolutionary environment.

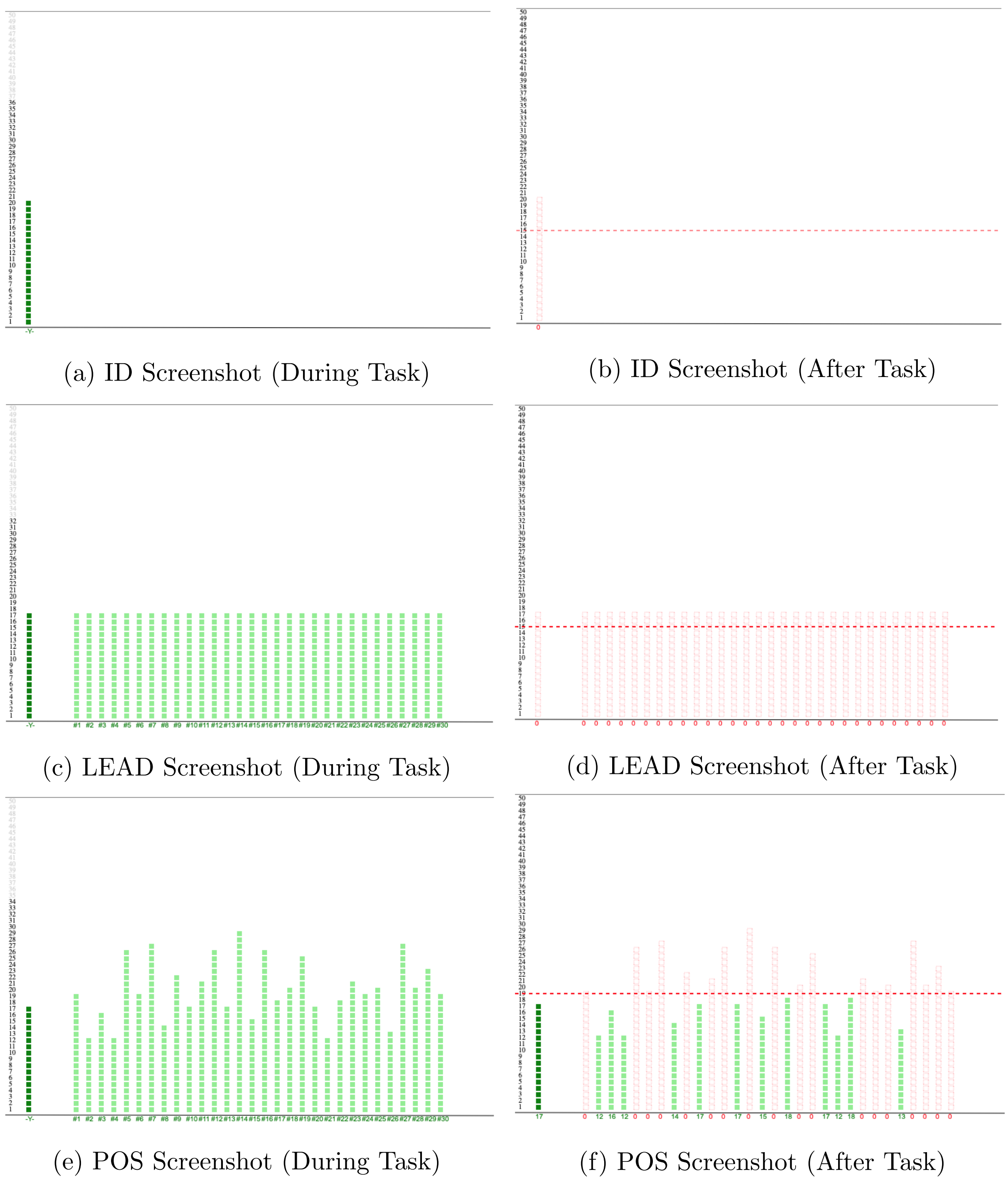
These challenges and questions are not unique to the ARH—they are generic to evolutionary models of economic behavior. Precisely *because* of the resulting ambiguity over the scope and reach of evolutionary hypotheses like the ARH, it is crucial to subject those hypotheses to stringent empirical examination. In Section 3, we propose an experiment to test the ARH and in Section 4 we show how this experiment produces a distinctive test of the ARH.

3 Experimental Design

Our experimental design tests the ARH by comparing measured risk preferences in four related tasks that are isomorphic to one another under standard economic theory, but are differentiated under the ARH. In treatment ID (idiosyncratic risk), subjects face lottery outcomes that are uncorrelated with outcomes for other subjects. In three additional treatments (LEAD, POS and NEG) subjects face lottery outcomes that are correlated in various ways with the payoffs of other subjects. The experiment is also designed to study subjects' preferences over exposure to these correlation structures. In this section we describe these experimental tasks in detail. In Section 4 we formally describe ten distinctive predictions the ARH makes for these tasks.

Each of our experimental treatments is built around a novel variation on the "Bomb Risk Elicitation Task" (BRET) task (Crosetto and Filippin, 2013), visualized in Figure 1a. When the task begins, the computer starts a clock that runs from 1 to 50 (visualized as darkening numbers on the left side of the screen). At each date of the clock, $t = 1, \ldots, 50$, a new green square automatically appears on the screen and is added to the vertical stack shown on the left side of the screen. When the subject

evolutionarily fittest type is consistent with maximization of Nature's utility, as given above in (1). Heller and Nehama (2023) instead proposes a dual utility approach where independent gambles are evaluated according to one utility function, while correlated gambles are evaluated according to another—more risk averse—utility function.



(a) ID Screenshot (During Task)

(b) ID Screenshot (After Task)

(c) LEAD Screenshot (During Task)

(d) LEAD Screenshot (After Task)

(e) POS Screenshot (During Task)

(f) POS Screenshot (After Task)

Figure 1: Screenshots from experimental software.

presses a button, the computer stops adding squares to the stack, though the clock continues to tick. After the clock reaches 50, a "laser" automatically fires from a uniform random vertical location between 1 and 50. If the laser strikes the subject's stack (as in Figure 1b), the stack is destroyed and the subject earns nothing for the task. If the laser misses the stack (e.g. fires above the subject's stack), she earns \$0.10 for each box in her stack.

The task provides an intuitive trade-off for the subject between risk and return. A risk neutral agent will stack $b = 25$ boxes; a risk averse subject will stack $b < 25$ and a risk seeking subject will stack $b > 25$. The coefficient of relative risk aversion from the CRRA utility function, $v(x) = x^{\mu}$, can be calculated from the subject's choice to stack b boxes as $\mu = \frac{b}{50-b}$. More generally, if subject r is more risk averse than subject s, in the sense of having a utility function with a globally higher Arrow-Pratt risk aversion coefficient, then r will choose a lower box—the less risky choice—than will s (Lemma **??** in the Online Appendix).

3.1 Varying Aggregate Risk

The core of the experiment is a series of variations on this basic task that vary the degree and kind of aggregate risk subjects are exposed to.

ID Task: At one extreme, is our **idiosyncratic** task labeled **ID**, the variation of the task described in the previous subsection (and visualized in Figure 1a and 1b). In this version of the task, the subject is making her choice "alone" without any reason to believe her payoffs are correlated with those of any other subject.[18]

LEAD Task: At the other extreme, Figure 1c shows a second variation on the task that we call **leader** (**LEAD**). In this version of the task, the subject is grouped with 30 other subjects and chooses a level of risk exposure for the entire group. Along the horizontal edge of the screen are a series of numbers, each representing a different subject from the group. At each of these positions, each subject has a different stack

[18] To the degree that subjects independently form a belief that their earnings *are* correlated with those of other participants in the experiment, it will simply weaken our main treatment effects. Thus, to the degree this is a concern, it suggests our results represent a lower bound on the effects we are measuring.

of boxes. To the far left is the subject's own stack. The task unfolds exactly as in **ID**, except that in each tick of the clock a new green box is added to *each* of the 31 subjects' stacks. The subject decides not only when to stop growing her own stack—her decision applies for each of the other 30 subjects as well. If the laser strikes *any* subject's *stack*, it will strike all subjects' stacks and all subjects will earn $0 for the task (as visualized in 1d); if the laser fires from above the subject's stack, all subjects will earn $0.10 \times b$. This treatment induces perfect positive correlation in payoffs between all members of the group.

POS Task: Figure 1e shows a third variation that we call **positive correlation** (**POS**). Unlike in LEAD, in POS subjects' choices have *no impact* on outcomes for other subjects. Instead, the subject is shown actual box choices made by 30 other subjects (making the presence of others highly salient and experimentally fixing subjects' beliefs), and must make a decision of how large of a stack to produce for herself given this information. In making this decision, the subject is aware that the *same laser* will be drawn at the same location for all 31 subjects and may strike some subjects' stacks but not others, depending on their levels of exposure (as in Figure 1f). This treatment produces a positive correlation in risk exposure across subjects.

NEG Task: Finally, we ran a fourth variation that we call **negative correlation** (**NEG**) that serves as a kind of placebo and a control for the POS task. Just as in POS, the subject is shown box choices from 30 other subjects, and knows all 31 members of the group will face a laser firing from the same random, vertical position. In fact, subjects in NEG are matched with the exact same other subjects and therefore are shown the exact same information, making it perfect control for non-ARH effects of information on others' choices in POS. However, unlike in POS, the subject's stack is "hung from the ceiling": instead of accumulating from the bottom of the screen, it extends progressively downward from the top of the screen. Since her 30 counterparts' boxes are stacked from the bottom of the screen, and all subjects face a laser firing from the same vertical location, this treatment induces negative correlation in payments between the subject's own payment and the payment of other subjects. The NEG Task looks identical to the POS Task visualized in Figures 1e and 1f, except that the subject's own stack (shown on the left side of the screen) extends from the *top* of the screen rather than the bottom.[19]

3.2 Treatment Design

The experiment consists of three treatments: LEAD, POS and NEG, each centered around a different, correspondingly named Task. In each treatment, each subject faces three tasks in sequence:

1. **ID:** The subject faces a single idiosyncratic Task.
2. **LEAD/POS/NEG**: The subject faces a single LEAD, POS or NEG Task, depending on the treatment.
3. **Choice:** The subject makes a choice of whether to face an additional ID or LEAD/POS/NEG Task (depending on the treatment). She then makes a decision in this third, self-selected task.

Importantly we **randomize the order** of Parts 1 and 2: roughly half of subjects face ID before LEAD/POS/NEG, and vice versa. This design allows us to measure the effect of aggregate risk on preferences for each individual subject and the randomized order of the tasks prevents this effect from being confounded with learning or task-to-task contamination. We randomly select only one of the three tasks for payment, and this is known ex-ante by subjects.

The experiment was conducted in two waves. In Wave 1 we collected data for the LEAD treatment. In the LEAD Task component of the treatment, all subjects made a decision for their 31 person group and were told that one random group member's choice would be applied to the entire group if the task was selected for payment.

In Wave 2, held immediately after Wave 1, we simultaneously ran the POS and NEG treatments, each using different subjects. Subjects in Wave 2 were matched (during their POS and NEG tasks) with a group of 30 subjects from the LEAD treatment. Since there is no payoff interaction in the POS/NEG treatment, this involved simply

[19]The NEG task is asymmetric, since only one subject has a negatively correlated payoff, while the 30 other subjects' payoffs are distributed exactly as in POS.

showing 30 ID choices from a sample of LEAD subjects, to POS and NEG subjects during their POS and NEG choices.[20] In order to generate the correlations described above, a single laser draw was made for all LEAD subjects during their ID task and for POS/NEG subjects during their POS/NEG tasks.[21]

3.3 Implementation

We conducted the experiment using custom software written by the authors in Javascript and deployed using Qualtrics. 361 subjects from the United States were recruited on the Prolific platform on December 17-18, 2020. Subjects were allowed to participate in no more than one treatment. The experiment took about 15 minutes on average and subjects earned, on average, just under \$5.00 including a fixed participation fee. We collected no pilot data for the experiment and pre-planned and prepared all of our treatments prior to data collection.

The experiment opened with instructions (reproduced in Online Appendix C) including a series of comprehension quizzes. Subjects showed evidence of high comprehension, with the modal subject making no quiz mistakes and over 75% making no more than one mistake. As is common practice with online experiments, we made an exante decision to remove subjects who made more than one comprehension mistake; our main hypotheses tests are unchanged if we instead include the entire sample. We also followed Crosetto and Filippin (2013) in removing any subjects from the dataset who chose $b = 0$ or $b = 50$ in any of their decisions as these decisions guarantee the subject a payoff of zero and therefore are a strong indication of confusion. This leaves us a sample of 282 subjects for our analysis.

Prior to each of the ID and LEAD/POS/NEG tasks, subjects were allowed two practice periods. Following each practice choice subjects were shown a series of example lasers to give them a feel of the range of possible outcomes as a function of their

[20] We selected the 30 LEAD subjects in order to (i) roughly match the ID mean from this treatment and (ii) to have relatively low variance to make it easier to interpret POS/NEG results.

[21] Subjects were not aware that the same laser draw was used across ID subjects in LEAD or in POS/NEG. If they had known this, we would expect the ID treatment to contain elements of aggregate risk. Importantly, to the degree subjects (somehow) infer this on their own, it will have the effect of *weakening* our main treatment effects. To the degree this is a concern, our results represent a lower bound on the effect we are measuring.

choice. After making their ID and LEAD/POS/NEG choices, subjects were shown a screen in which they were allowed to choose between an ID Task or a LEAD Task (or POS or NEG Task, depending on treatment) for their final task. When this third task (which we will call the "choice task") was complete, subjects completed a short demographic survey.

3.4 Understanding the Design

We designed this experiment to achieve several inferential goals. In Section 4 we formally show how the design provides a demanding test of the ARH. In this subsection we explain why we made the design choices we did and how they serve the goals of the experiment.

First, we wanted to use a risk elicitation task that is visual and intuitive to facilitate high comprehension in our online sample. To do this we adapted the BRET task (Crosetto and Filippin, 2013) which allows for fast and intuitive elicitation of the CRRA parameter. The high comprehension rate of our subjects suggests that we were successful in developing an intuitive task. Framing the BRET task as an investment problem allowed us, in aggregate treatments, to keep the problem unchanged except by adding the stacks of other participants horizontally alongside the subject's own stack. By describing the earnings risk as a laser firing from a random vertical position, we are able to make the correlation in risk intuitively clear in the aggregate risk tasks without changing anything else about the problem and without having to technically explain correlations.

Second, we wanted to gather within-subject evidence on the contrast between responses to idiosyncratic risk relative to aggregate risk. That is, we wanted to gather evidence on how subjects *change* their behavior in one setting relative to the other. We believe this provides particularly strong evidence of sensitivity to aggregate risk (if any emerges) and produces much stronger statistical power, particularly given well-documented heterogeneity in risk attitudes in typical subject pools. For this reason, every subject faced both the idiosyncratic task, ID, and an aggregate task. In order to avoid confounds due to learning or changes in risk posture with experience, we randomized the order in which subjects faced these tasks.

Third, as we discuss in more depth in Section 5.4 we wanted to vary the nature of correlations across aggregate tasks in such a way as to *create a distinctive empirical fingerprint*, allowing us to rule out confounding explanations for our results as a whole (e.g., social preferences, conformism etc.). Our LEAD treatment creates a strong (indeed, perfect) positive correlation in risks. However it also requires subjects to make payoff relevant decisions on other subjects' behalf. Although standard social preferences like inequity aversion, altruism and efficiency preferences shouldn't be relevant in this task (all subjects earn an identical amount), the responsibility for others' payoffs is an additional change in this treatment relative to ID above and beyond the introduction of positive payoff correlation.

Our POS treatment avoids this potential confound by inducing positive correlation in payoffs, while preventing the subject's actions from having any effect on others' payoffs, shutting down social preference explanations. Although it is in principle possible to run the POS treatment without showing subjects previous participants' choices, we made the important design decision to show subjects the choices of their counterparts in this treatment for two reasons. First, we thought it was important to make the presence of other subjects who share the same risk highly *salient* to subjects in order to increase our chances of observing possibly subtle evolutionary effects. Visually representing other subjects seems significantly more salient (and more credible) than simply verbally describing their presence. Second, showing this information allows us to control for subjects' otherwise heterogeneous beliefs about other subjects' choices, which seemed likely to significantly reduce noise in our elicitations.

This design decision, however, may introduce a different confound. Subjects in POS are given *information* on other subjects' choices and this may conceivably influence choices. That is, subjects may avoid risk in POS for heuristic (perhaps imitative) reasons unrelated to the positive correlation in payoffs hypothesized to induce an effect in Section 2. For this reason, we also ran the NEG treatment which provides subjects with the *exact* same visual environment and information on other subjects' choices as POS (we even match NEG treatment subjects with the same set of 30 counterparts to make this informational control *perfect*), but removes the positive payoff correlation hypothesized to generate a treatment effect in Section 2. Under the hypotheses discussed in Section 4, in this treatment (unlike POS and LEAD) subjects should not become more risk averse relative to settings of idiosyncratic risk, making it an ideal control.

Finally, in order to make our test of the ARH even more stringent we wanted to include a task that measures aggregate correlation aversion (ACA) directly by measurings subjects' *preferences* for facing correlated versus uncorrelated risk on the extensive margin. Thus, we included a third task in the experimental design, that is directly *selected* by the subject. In particular at the end of the experiment, each subject was allowed to choose to either face a task alone (idiosyncratic risk) or grouped with 30 other subject (aggregate, correlated risk). We have little interest in subjects' behavior in this third task since self-selection into this task raises endogeneity concerns and we therefore do not include this data in our analysis. Rather we include the task simply in order to elicit subjects' preferences over payoff correlations in an incentive compatible way, allowing us to test additional distinctive implications of the ARH that are not easily explained by other mechanisms.[22]

4 Hypotheses

We now describe a set of empirical hypotheses, derived from the ARH, that our experiment is designed to test. As we show in the Online Appendix, all of our hypotheses can be derived from a direct application of Robson (1996) (as sketched in Section 2) to payoff settings that mirror those of our four treatments (ID, LEAD, POS

[22]To be precise, this task elicits a *weak* preference for idiosyncratic relative to aggregate risk. We considered attaching costs to selection in order to make the elicited preferences strict, but decided that this would add too much complication to the design. As we show below, subjects' decision in this task is nonetheless highly predictive of attenuated risk aversion along the intensive margin, suggesting that most of our subjects express strict preferences.

and NEG) and our Choice tasks in the large population limit.[23,24] The experiment is designed around an unusually large number of hypotheses (ten in total). This is deliberate in that our goal was to design an experiment in which the ARH produces a rich empirical *fingerprint*: a pattern of observed behaviors that collectively bear the distinctive stamp of the model. Doing this has two benefits. First, the more distinctive and idiosyncratic the fingerprint is, the less likely it is to be confused with other ultimate or proximal accounts of behavior. Second, the richer the fingerprint, the more behavioral discoveries the evolutionary model will allow us to uncover. Indeed, as we emphasize in the concluding Discussion, the primary motivation for testing an evolutionary theory is not to evaluate the merits of the story, per se, but to use the theory for discovering behaviors we might not otherwise find using more conventional approaches to generating behavioral hypotheses.

Our first hypotheses focus on the implications of the ARH for risk taking in our tasks. Let b_r denote the size of a subject's stack in task $r \in \{\text{ID, POS, LEAD, NEG}\}$ of the experiment. Recall that lower b_r implies greater risk aversion. First, our hypothesis about choices in LEAD:

Hypothesis 1 *Subjects will make less risky choices in LEAD than in ID:* $b_{LEAD} < b_{ID}$.

An identical result applies to our POS task, giving us a second hypothesis that mirrors the first:

Hypothesis 2 *Subjects will make less risky choices in POS than in ID:* $b_{POS} < b_{ID}$.

As we show in Online Appendix in Section **??**, this Hypothesis and the next are consequences of the joint influence of aggregate correlation aversion (ACA) and IAIR, two key implications of the ARH discussed in Section 2.1; if subjects were instead

[23] Of course the question of whether decision makers obey predictions derived for large populations in smaller scale settings like organizations or lab cohorts is one of the main motivations for our experiment. It is difficult to obtain definite theoretical predictions for finite populations; however, the propositions derived above in Section 2.1 will hold approximately for sufficiently large but finite populations.

[24] Some of our Hypotheses (e.g., 2, 3 and 4) can also be derived under less direct conditions (e.g., if preferences respect ACA and IAIR).

correlation neutral, then this would imply $b_{ID} = b_{POS} = b_{LEAD}$, and if they were correlation loving, the above inequalities would reverse.

Our third hypothesis is that this effect can be shut down simply by reversing the correlation in the subjects' payments, as we do in the NEG treatment. This provides a kind of placebo test for the first two effects, ruling out some alternative explanations for these results (as we discuss in Section 5.4):

Hypothesis 3 *Subject will not make less risky choices in NEG than in ID, but will instead make more risky choices:* $b_{ID} < b_{NEG}$.

As discussed, Hypotheses 2 and 3 rely on IAIR, a consequence of the ARH. If IAIR failed to hold such that the subject had *intensified aversion to risk in own payoffs when others' payoffs are higher*, then it is possible for Hypothesis 2 and Hypothesis 3 to hold even for decision makers with an affinity for correlated payoffs rather than an aversion. With this in mind, we add an additional, stronger, hypothesis that holds *if and only if* the subject is aggregate correlation averse as specified by the ARH (Proposition **??** in the Online Appendix):

Hypothesis 4 *Controlling for ID risky choices, subjects will tend to be more risk averse in POS than they are in NEG:* $b_{POS} < b_{NEG}$.[25]

The Choice task, in which subjects make a direct choice of whether to make their final decision in LEAD/POS/NEG or ID, allows us to test subjects' distaste for correlations (ACA) more directly. As we show in the Online Appendix (Section **??**), because agents in Robson's model are aggregate correlation averse, they should be expected to display the following regularities in our Choice tasks:

Hypothesis 5 *Subjects will prefer ID to LEAD tasks: subjects tend to choose ID in the Choice task in the LEAD treatment.*

Hypothesis 6 *Subjects will prefer ID to POS tasks: subjects tend to choose ID in the Choice task in the POS treatment.*

[25] In order to test for Hypothesis 4, we control for ID choices by comparing the average POS and NEG choices for subjects that made the same choice in their ID tasks.

Hypothesis 7 *Subjects will prefer ID to NEG tasks: subjects tend to choose ID in the Choice task in the NEG treatment.*[26]

A final set of hypothesis is intended to further establish the ARH as driving the results of our tests of Hypotheses 1-7. In particular, if the ARH is behind failures to reject Hypotheses 1-3, then the effect of correlation on risky choices in Hypotheses 1-3 will be strongest for subjects that choose in line with Hypotheses 5-7. We thus hypothesize:

Hypothesis 8 *In the LEAD treatment, risk aversion relative to choice in ID will increase more strongly among subjects that choose ID in the Choice task than among those that choose LEAD.*

Hypothesis 9 *In the POS treatment, risk aversion relative to choice in ID will increase more strongly among subjects that choose ID in the Choice task than among those that choose POS.*

Hypothesis 10 *In the NEG treatment, risk aversion relative to choice in ID will decrease more strongly among subjects that choose ID in the Choice task than among those that choose NEG.*

Finally, as discussed in Section 2.2, by explicitly testing these hypotheses we are implicitly testing more fundamental hypotheses that are not direct implications of the theory. Most importantly by testing these hypotheses on relatively small experimental cohorts we are implicitly testing the hypothesis that Nature implemented the ARH in a relatively insensitive fashion so that individuals are sensitive to correlated risks even in relatively small groups (rather than being sensitive only to aggregate risks that affect the entire genetic population—as a literal reading of the model might suggest). This is a hypothesis on which the evolutionary theory is silent, but which is important for understanding the ARH's relevance to economics.

[26]Hypotheses 5-6 are immediate implications of the evolutionary theory underlying the ARH. Specifically, in an evolutionary model with aggregate risk, where consumption is determined as in our experimental tasks, the ID environment yields a higher population growth rate than the POS and LEAD environments. For Hypothesis 7 the same is true for a sufficiently sized population. The POS and NEG environments are identical, except for the payoffs of the subject. Hence, when the population is large, the two environments yield the same population growth rate.

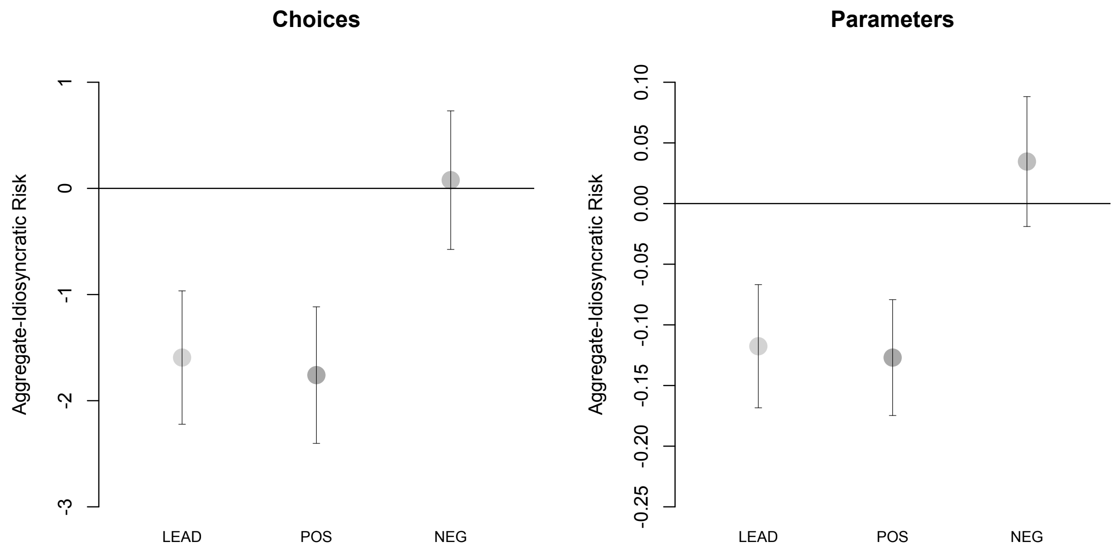


Figure 2: The mean within-subject change in risk taking for aggregate risk tasks versus idiosyncratic risk tasks. *Notes: Error bars are standard errors. The left panel shows the mean choice difference while the right panel shows the mean change in the estimated CRRA parameter.*

5 Results

The left panel of Figure 2 summarizes our main results by plotting the mean *within-subject difference* in the number of boxes "stacked" in the idiosyncratic risk (ID) treatment vs. each of the aggregate risk treatments (LEAD/POS/NEG). The right panel shows the mean within-subject difference in estimated CRRA parameters. The results show that, as predicted by the ARH, subjects are significantly more risk averse under aggregate than idiosyncratic risk in the POS and LEAD but *not* in the placebo NEG treatment. Because this is a within-subject comparison, it automatically differences out subject-level variation in risk attitudes, producing a particularly well-powered test of aggregate risk aversion.[27] In Section 5.2 we show that this strong evidence in favor of the ARH holds statistically using several approaches.

Figure 3 and Table 1 break down this data by choice to allow for a complemen-

[27]Wilcoxon tests performed at the treatment level indicate that the order with which subjects were assigned the two tasks has no discernible effect on the size of these differences ($p = 0.662$, 0.475 and 0.361 for the LEAD, NEG and POS treatments, respectively).

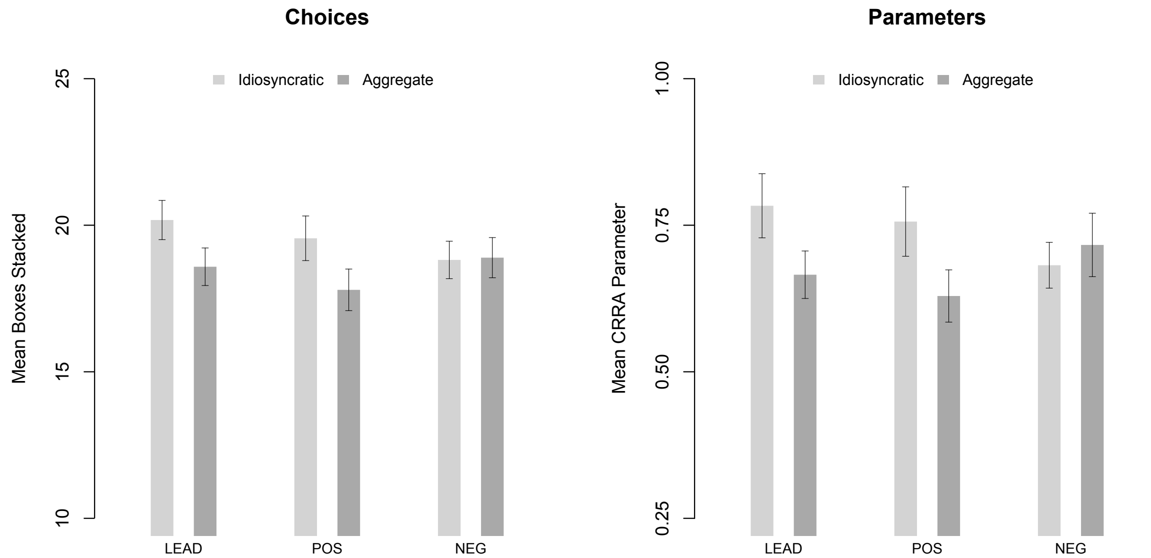


Figure 3: Mean risk taking in idiosyncratic vs. aggregate risk tasks, by treatment. *Notes: Error bars are standard errors. The left panel shows the mean choices while the right panel shows mean estimated CRRA parameters.*

tary (though lower powered) between-subjects comparison. Figure 3 plots raw means (with standard error bars) for the idiosyncratic risk task (ID), and the aggregate risk task (LEAD, POS or NEG, depending on treatment) for each of our three treatments. Again, the left hand panel shows the mean choice while the right hand panel shows the mean CRRA parameter estimate for these choices. Table 1 complements Figure 3 statistically by reporting a series of simple OLS regressions (with standard errors clustered at the subject level) that summarize the data numerically. The table includes three specifications (1, 2 and 3) for each of the three treatments (LEAD, POS and NEG). In each treatment, specification (1) includes a dummy variable for the aggregate task (LEAD, POS or NEG, depending on treatment) that measures the difference between aggregate and idiosyncratic risk-taking. In specification (2) we add an additional dummy variable that takes a value of 1 if the aggregate task was presented to the subject second (to examine order effects). Specification (3) repeats specification (2) but includes only the (majority of) subjects that show evidence of correlation aversion (i.e. who chose the idiosyncratic risk task rather than the treatment's aggregate risk task in the third part of the experiment). Table 2 in Online

Appendix B presents the same specifications for subject-wise parameter estimates rather than for for raw choices. In each of these regressions, the intercept term gives the estimated choice/CRRA parameter for subjects in the idiosyncratic risk (ID) Task.

5.1 Preliminaries: ID Behavior

The baseline ID Task was designed to measure aversion to risk in own payoff when outcomes are statistically independent across subjects, allowing for comparison with standard risk elicitation tasks from the prior literature (which focus almost exclusively on measuring risk preferences over population-independent gambles). The light gray bars in the left hand panel of Figure 3 shows the mean number of boxes the subjects stacked (left panel) and the resulting mean CRRA parameter estimates (right panel) for idiosyncratic (ID) Tasks in each of our treatments.

Across treatments, the mean subject stacked 19.5 boxes and the corresponding mean CRRA parameter across treatments is 0.738, revealing moderate risk aversion which is well within the normal range of idiosyncratic risk preferences estimated in the prior experimental literature. Although there is slight variation in the mean level of idiosyncratic risk aversion across treatments (particularly in the NEG treatment), these differences are not statistically significant: pairwise Wilcoxon tests fail to reject the hypothesis that they differ across treatments ($p > 0.2$ in all comparisons).

The *Second* dummy variable in specifications (2) and (3) of Table 1 and 2, meant to identify order effects, is uniformly indistinguishable from zero. As in past work with similar tasks, there is a slight overall tendency for subjects to bear more risk with experience, but the effect is not statistically significant. In any case, our randomization of the ordering of the idiosyncratic and aggregate tasks prevent any systematic confounds between such an effect and our main treatment effects.

	LEAD	LEAD	LEAD	POS	POS	POS	NEG	NEG	NEG
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Constant	20.177***	20.014***	20.308***	19.554***	19.164***	19.763***	18.816***	18.521***	19.333***
	(0.671)	(0.789)	(0.999)	(0.764)	(0.828)	(0.933)	(0.641)	(0.722)	(0.799)
Aggregate	−1.594**	−1.575**	−2.356***	−1.759***	−1.831***	−2.359***	0.078	0.115	−0.306
	(0.630)	(0.642)	(0.790)	(0.645)	(0.640)	(0.720)	(0.654)	(0.656)	(0.818)
Second		0.308	−0.781		0.853	0.159		0.552	0.667
		(0.642)	(0.790)		(0.640)	(0.720)		(0.656)	(0.818)
Observations	192	192	138	166	166	98	206	206	148
R^2	0.015	0.016	0.031	0.017	0.021	0.033	0.00003	0.002	0.004
Adjusted R^2	0.010	0.005	0.017	0.011	0.009	0.013	-0.005	-0.008	-0.010
Note:									*p<0.1; **p<0.05; ***p<0.01

Table 1: Estimated drivers of choices (number of boxes stacked). *Notes: Each column reports a separate specification run on data from the treatment (LEAD, POS and NEG) listed at the column's top. Constant measures mean behavior in the ID Task. Aggregate is a dummy variable for the aggregate task (LEAD, POS or NEG depending on treatment). Second is a dummy that takes a value of 1 if the Task was experienced by the subjects second in order. Specifications (1) and (2) make use of the entire dataset, while (3) makes use of data only from the (majority of) subjects who expressed a preference for the ID task over the aggregate task at the end of the experiment.*

5.2 Aggregate Risk Aversion

Our main question is whether (and under what circumstances) attitudes to aggregate risk differ from attitudes to idiosyncratic risk. We begin with the LEAD Task in which subjects choose a uniform level of risk for their entire 31 person cohort so that payoffs are *perfectly positively correlated* across the 31 participants. As discussed above, our sharpest evidence comes from the distribution of within-subject differences between ID and LEAD, visualized in Figure 2. In both panels, the dot for LEAD is far below zero, indicating that the average subject makes a more risk averse choice in LEAD than in ID. A Wilcoxon test indicates that this effect is highly significant ($p < 0.01$), with the implied CRRA parameter decreasing by 0.118 in LEAD relative to ID for the average subject. We find the same results in the between-subjects data visualized in Figure 3, which shows that subjects in LEAD (i) choose to expose themselves to less risk in the LEAD task than they do in the ID task (left hand panel) indicating (ii) stronger risk aversion (the right hand panel) than in ID. The *Aggregate* dummy in Tables 1 and 2 show that this difference is statistically significant, with the CRRA parameter falling 15% in the LEAD task relative to the ID task.

Result 1 *Subjects tolerate significantly less risk when choosing on behalf of a group than when choosing only for themselves. This is evidence in favor of Hypothesis 1.*

Can this result be interpreted as a change in subjects' risk preferences when payoffs are correlated, or is this merely a consequence of the fact that the subject must make decisions on behalf of others in the LEAD treatment? For example, perhaps the result is driven by interdependent risk preferences, where individuals feel uncomfortable making risky choices on behalf of others. To answer this question, we ran the POS treatment in which subjects do not influence other subjects' risk exposure at all, but instead are only informed that their payoffs are positively correlated with those of other members of their cohort. We find an almost identical pattern of results in the POS as we do in LEAD. Figure 2 shows, using our primary within-subjects analysis, that, as in LEAD, the dot for POS falls well-below zero, indicating signifantly more risk aversion in POS than ID. A paired Wilcoxon test confirms that this is a highly statistically significant effect ($p < 0.01$). Figures 3 and Tables 1 and 2 show the same in a between-subjects comparison, with estimated CRRA coefficients dropping by nearly 17%.

Result 2 *Subjects tolerate significantly less risk when their payoffs are positively correlated with those of others. This is evidence in favor of Hypothesis 2.*

Finally, it is possible that the results from POS are not driven by the positive correlation between the DM's and the others' payoffs, but rather by social learning or some other effect of being informed of other subjects' choices. To rule out such effects, we designed the NEG treatment, a kind of placebo treatment in which (as in POS) subjects observe other subjects' choices (indeed in our design the POS and NEG DM's observe the exact same distribution of others' choices). The only difference is that in NEG own payoff is negatively correlated with the payoffs of others, rather than positively correlated, as it is in POS.

In stark contrast to LEAD and POS, average risk taking in NEG is virtually identical to risk taking in ID. As is clear from Figure 2, the within-subject change in behavior in NEG relative to ID is in fact slightly *positive* and Wilcoxon tests confirm that the difference is indistinguishable from zero ($p = 0.821$). Once again, Figure 3 and Tables 1 and 2, provide complementary between-subjects evidence: mean choices are exactly the same and parameter estimates are even slightly *larger* in NEG than in

ID as suggested by Hypothesis 3. The *Aggregate* dummies in Tables 1 and 2 are statistically insignificant indicating no difference in behavior in NEG relative to ID.

Result 3 *Subjects do not behave differently when their payoffs are negatively correlated with those of others. This is evidence that partially supports Hypothesis 3.*

We interpret this last result as partial support for Hypothesis 3. As predicted, switching from positive to negative correlation eliminates the increase in observed risk aversion when going from ID to POS. This is important because it identifies positive correlation as the driving force behind the differences in risk taking that we see across the ID and aggregate treatments. The main purpose of this treatment is to serve as such a placebo, and in this sense our findings are strikingly consistent with the Hypotheses and, by extension, the ARH. But contrary to Hypothesis 3 we do not find evidence that it has the further effect of causing a *decrease* in risk aversion relative to ID.

Next, the ARH predicts (Hypothesis 4) that subjects will be more risk averse in POS than in NEG, when controlling for ID choices. Unlike Hypotheses 1-3, this is a special prediction of correlation aversion and does not rely on IAIR, or a large population. This prediction thus more strongly implicates correlation aversion than Hypotheses 1-3 do. We study this by controlling for ID behavior in two ways. First, we look at deviations of POS vs. NEG from ID, as plotted in Figure 2. Examination of this Figure shows evidence of much stronger risk aversion in POS than NEG (relative to ID choice) and a Wilcoxon test allows us to reject the hypothesis that they are the same ($p = 0.02$). Second, calculating the mean POS and NEG choices corresponding to each observed ID choice, we find that POS choices are more risk averse than NEG choices in 2/3 of cases.

Result 4 *Subjects are more risk averse in POS than in NEG, when controlling for ID risk preferences. This is evidence in favor of Hypothesis 4.*

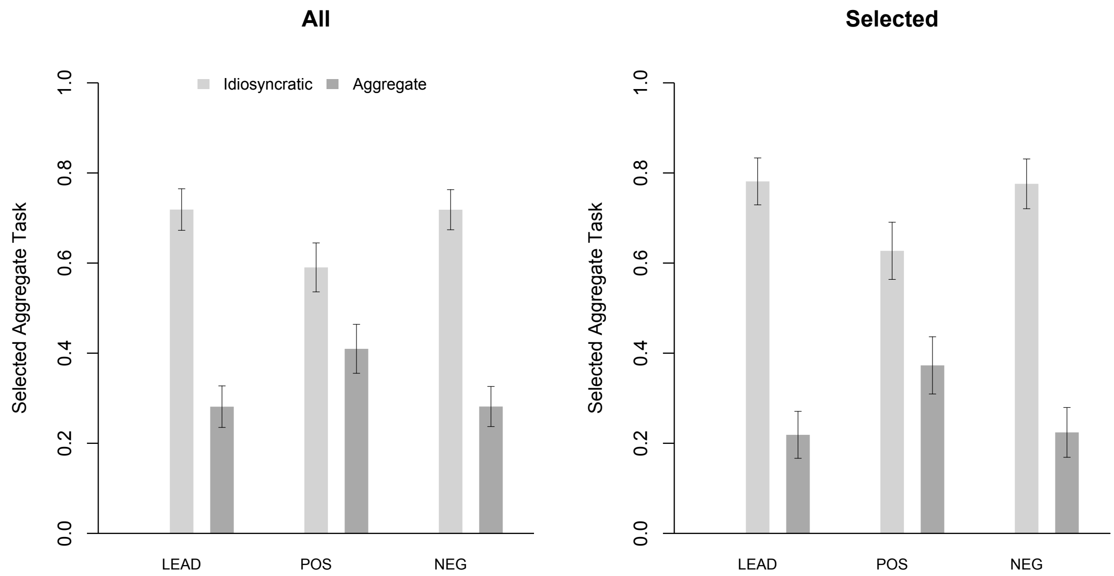


Figure 4: Rates at which subjects chose idiosyncratic (light gray) versus aggregate (dark gray) payoff settings for their final Task. *Notes: The left hand panel shows results for the entire sample, while the right hand panel shows results for subjects who are not less risk averse in POS and LEAD, and not more risk averse in NEG.*

5.3 Correlation Aversion

Our first four findings are consistent with Aggregate Risk Aversion (ARA), a key prediction of the ARH. The final part of our experimental design examines further evidence of the ARH, using a revealed preference approach, by studying whether subjects make extensive margin choices in favor of environments involving risks that are independent across individuals. After experiencing both an ID and an aggregate (POS, NEG or LEAD, depending on treatment) Task, we allowed subjects to choose between the two types of risk for their third and final task. Under the hypothesis of aggregate correlation aversion, we expect subjects to tend to choose ID over LEAD/POS/NEG (Hypotheses 5-7).

Over all three treatments, subjects are nearly 70% likely to choose ID rather than LEAD/POS/NEG, and are significantly more likely to choose ID than a correlated treatment by a proportions test ($p < 0.0001$) Figure 4 breaks this down by treatment, showing the rate at which subjects choose ID over the aggregate Task (in light gray) and the rate at which they choose the reverse (in dark gray). In both LEAD and NEG, 72% of subjects choose ID; curiously, the effect is somewhat smaller in POS than in the other two treatments with a smaller majority (about 60% of subjects) choosing ID.[28]

Result 5 *Overall, subjects are more than twice as likely to choose the idiosyncratic task than the aggregate tasks. This is evidence in favor of Hypotheses 5-7.*

Finally, the ARH predicts a tight linkage between aggregate correlation aversion (ACA) and aggregate risk aversion (ARA), because it hypothesizes that the latter is a consequence of the former (i.e. that the increase in risk aversion is due to a distaste for correlation). To test this directly, we re-examine within-subject differences between LEAD/POS/NEG and ID, but break the data down into subjects who selected ID in Part 3 of the experiment (evincing aggregate correlation aversion) and those who did not. In both POS and LEAD, we continue to find strong within-subject evidence of an increase in risk aversion relative to ID among aggregate correlation averse subjects ($p < 0.001$ in both case by Wilcoxon tests). However we find no such evidence among the subjects who do not show evidence of aggregate correlation aversion ($p = 0.377$ and $p = 0.759$ in POS and LEAD respectively). Finally we find that the magnitude of the increase in risk aversion (the difference between POS/LEAD and ID behavior) is significantly larger among aggregate correlation averse subjects than other subjects ($p < 0.0001$). The evidence therefore shows that risk choices are connected to correlation attitudes in exactly the manner predicted by the ARH.

Result 6 *We only find evidence of Hypotheses 1 and 2 only among subjects that reveal correlation aversion in the choice task. This is evidence in favor of Hypotheses 8 and 9.*

We also find partial evidence for Hypothesis 10. Unlike in LEAD and POS, we find no evidence of increased risk aversion among subjects that choose the ID task over

[28] These results strengthen for the subset of subjects who show evidence of the ARH by complying with Hypotheses 1-3; in this sample we can reject the hypothesis that subjects LEAD/POS/NEG as often as ID for each treatment individually at the 10 percent level ($p < 0.07$ in each case).

NEG. This is strongly consistent with the predictions of the ARH as summarized in Hypothesis 10. However, as with Result 3, we find no evidence of less risk aversion among subjects that choose ID over NEG. Just as with Result 3, we therefore find that the NEG treatment serves its function as a placebo, but we do not find evidence that it induces more risk loving behavior.

5.4 The Fingerprint of the ARH

We designed the experiment to test an unusually large set of hypotheses in order to produce a distinctive empirical fingerprint of the ARH. Collectively this fingerprint seems difficult to confuse with evidence of behavioral hypotheses other than the ARH. Although any given subset of these hypotheses might hold for some alternative theory, we believe it is difficult for any one alternative to generate the entirety of our results. Thus, to the degree that we have built our experiment around a sufficiently rich set of hypotheses, the ARH offers a parsimonious explanation for our results that few other behavioral theories can deliver.

All ten of the hypotheses of the ARH that we articulated in Section 4 are borne out in the data, usually completely and in two cases partially. As the ARH predicts, subjects become significantly more risk averse when payoffs are positively correlated (Hypotheses 1-2) but this effect disappears in our placebo NEG treatment in which risks are negatively instead of positively correlated (Hypothesis 3). Subjects also show strong evidence of being averse to aggregate correlations (Hypotheses 5-7), a key implication of the ARH, strongly linked to aggregate risk aversion. Indeed, aggregate risk aversion is strongly predicted by aggregate correlation aversion in the LEAD and POS treatments (Hypotheses 8 and 9) but not in our placebo NEG treatment (Hypothesis 10), confirming the link between the two behaviors stipulated by the ARH. In short, our results almost perfectly match the distinctive fingerprint of the ARH. The only exception is that our placebo NEG treatment, while successful at shutting down aggregate risk aversion (as the ARH predicts), does not lead to significantly more risk loving behavior as expected as described in Hypotheses 3 and 10.

Because our set of predictions is extremely rich, it is difficult to account for their success using alternative theories. To show this, we consider several salient alternative explanations.

Expected Utility Theory: Standard expected utility theory is starkly inconsistent with Hypotheses 1-4. It also does not predict in any direct way Hypotheses 5-10.

Altruism: We find that subjects display more risk aversion in the LEAD task than in the ID task (Hypothesis 1), which is consistent with the ARH since the treatment induces positive payoff correlations. But since the treatment involves direct choices over others' payments, an alternative explanation is that it is an outgrowth of some sort of direct social preference. For example, if subjects believe that some of their counterparts are more risk averse than they are, they might make less risky choices out of a sense of altruism (e.g., in order to match the risk preferences of the group). Likewise, altruistic DM's might prefer the ID tasks over LEAD in the Choice task if they believe that ID would give the other subjects the flexibility to make their own choices. However, this kind of altruism does not explain our support for Hypotheses 2-4, 6-7 or 9-10 since subjects have no control over the payoffs of others in the POS and NEG treatments.

Social Learning: We find that subjects display more risk aversion in the POS treatment than in the ID treatment (Hypothesis 2) which the ARH predicts because payoffs are positively correlated in POS. But an alternative explanation for this behavior might be some form of social learning. Subjects after all, in this treatment, get to see the choices of 30 other subjects which might influence their own choices. But this explanation for choices in POS seems difficult to square with Hypotheses 3 and 4: we deliberately give subjects in the NEG treatment the exact same social information as subjects in POS and fail to find more risk aversion in NEG than in ID, as is found in the POS treatment. Likewise, social learning does not predict Hypothesis 1 because there is no social information in LEAD to drive the observed increased risk aversion observed in LEAD. Moreover, social learning would seem to suggest the opposite of Hypothesis 5: subjects keen to imitate others might be expected to instead prefer the POS treatment to the ID treatment.

Relative Earnings Preferences: Similarly, increased risk aversion in POS (Hypothesis 2) might be a consequence of a desire on the part of subjects to earn relatively higher amounts than their counterparts by increasing the chances they will earn positive amounts when the others earn zero. This might also be consistent with Hypothesis 3. But such "status" preferences do not predict Hypothesis 1 for behavior in LEAD where all subjects earn identical payoffs.

Increasing Risk Aversion: As discussed above in Section 4, an alternative to the ARH can yield Hypotheses 1-3. Specifically, if subjects' aversion to risk in own payoff is increasing in other subjects' earnings (for whatever reason) the behaviors described by Hypotheses 1-3 might be observed even for a DM with an affinity for correlation in own and others' payoffs. However, as we show in the Online Appendix, correlation affinity rules out Hypothesis 4: that people will be more risk averse in POS than in NEG. Furthermore, the results from the Choice tasks further help to rule out correlation affinity by showing that subjects have, to the contrary, direct *aversion* to payoff correlations.

Do these alternative behavioral stories exhaust the set of proximal explanations for some of our hypotheses? Almost certainly not. However, to the extent our hypotheses have produced a sufficiently rich and idiosyncratic fingerprint, few alternative explanations will as parsimoniously account for all of them as does the ARH. Of course to the extent we have failed, the richness of evolutionary models allow us to enrich our set of hypotheses further to distinguish the ARH from alternative stories. In which case, the model will generate new opportunities for behavioral discovery.

6 Discussion: Evolution as a Source of Behavioral Hypotheses in Economics

Our results suggest that human preferences for risk are affected by the cross-population correlation structure of those risks. A long line of theoretical work in evolutionary economics predicts this fundamental departure from standard theory (the "aggregate risk hypothesis" or the ARH), and we find strong empirical evidence to support it. Indeed, we subject the ARH to a demanding test and show that our data matches a large and distinctive collection of predictions made by this evolutionary theory.

These findings are of potential importance, first, because they so sharply contrast with standard economic reasoning and methodological practice. A pervasive idea in economics is that we can use the tastes and preferences measured in one context (e.g. consumer choice or direct preference elicitation) and project them onto other contexts (e.g. trade in equity markets) to make predictions, estimate parameters, produce calibrations, form counterfactuals and make welfare judgments. Our findings suggest that, at least when it comes to risk preferences, this standard approach can produce severely misleading results. Indeed, the wedge between idiosyncratic and aggregate risk preferences predicted by the ARH offers a promising explanation for open mysteries like the famous equity premium puzzle (Robson and Orr, 2021), which illustrates this wedge with particular clarity. Because these revealed preference extrapolations are fundamental to the practice of economics, employed in every sub-field, the implications of findings like these for the field seem potentially significant.

More positively, our findings also raise the possibility that seemingly non-social determinants of behavior, such as risk-preferences, can be affected by social aggregation. This opens up potential new vistas in behavioral economics for modeling the psychologically distinctive character of aggregate interactions. For instance, it may be that one of the costs (or benefits) of aggregate social interaction is that it induces caution and conservatism by diminishing decision makers' risk tolerance. This may, for example, help to explain the often-noticed conservatism of large organizations relative to more nimble and exploratory upstarts. More broadly, the possibility that *other* aspects of behavior may similarly take on a special character in aggregate interaction introduces new avenues of investigation for behavioral economists. Exploring the special psychology of aggregate interactions is a task behavioral economists seem especially well suited for (relative to some other fields studying human psychology), given that economics is historically an aggregate science.

Finally (and more broadly), our work suggests that there are large potential gains from taking the predictions of evolutionary economic models seriously as *behavioral hypotheses* in economics. In the last several decades, a significant body of economic theory (the "evolutionary foundations" literature) has emerged modeling how the *proximal* explanations for behavior in economic models (like preferences, beliefs and reasoning capacities) might have been shaped by the *ultimate* explanatory forces of evolution. To date, findings from this rich literature have rarely been adopted for empirical testing or use in economic explanation, even in behavioral economics. Instead, economists have traditionally culled their behavioral hypotheses from normatively appealing axioms, casual empiricism, and, more recently (in, e.g., behavioral economics), from findings in other fields like psychology. Our findings suggest that there is value to expanding the reservoir of behavioral hypotheses to include evolutionary models.

Indeed, our investigation is illustrative of the potential upside of taking evolutionary models seriously as an additional source of behavioral hypotheses. The ARH makes such strange and distinctive predictions that it is hard to imagine deriving them from normative axioms, casual empiricism or findings from disciplines less focused on aggregate interaction, like psychology. As a result, we may never have thought to test the hypothesis that risk preferences change in the face of aggregate risk if not for the suggestion that this might be the case made by formal evolutionary economic models. Because evolutionary models account for different constraints and explanatory factors than do our typical methods for generating hypotheses, they have significant potential to suggest avenues for empirical investigation we otherwise would be unlikely to explore.

Of course, there may be reasons to question whether the behaviors implied by any given model of evolution should be wholesale adopted into the practice of economics. The partial equilibrium nature of typical evolutionary models, questions about their extensibility from the evolutionary environment to modern contexts, and ambiguities about what types of behaviors natural selection may have selected to satisfy evolutionary constraints all limit their immediate application. However, this is only to say that evolutionary hypotheses (like any other behavioral hypotheses) require empirical testing before they can be adopted to explain contemporary economic behavior. Our claim is not that evolutionary models should be taken at face value, but rather that they are a rich source of behavioral hypotheses that economists can draw from, producing predictions that our current reservoirs of hypotheses would not otherwise deliver.

A perennial criticism of the use of evolutionary explanations of behavior in the social sciences is that they are often "just so stories" (Gould, 1978), generating unfalsifiable explanations of well-known social phenomena. Certainly, this is a danger of the undisciplined use of post hoc evolutionary explanations, and pop-evolutionists have undoubtedly at times indulged in this style of explanation. But this criticism misses an important use of evolutionary models in economics and the other human sciences. It is not the direct evolutionary explanation for a phenomenon that produces useful hypotheses for economics, but rather the indirect, secondary and tertiary implications of that evolutionary explanation. It is not the phenomena the evolutionary model was intended to explain, but the phenomena the model unintentionally suggests—the excess explanatory power of the model—that is of value to behavioral economics.

The ARH is, in fact, an ideal case-in-point. When Arthur Robson built the first evolutionary model that yielded the ARH, he was not attempting to explain previously acknowledged differences between idiosyncratic and aggregate risk-taking.[29] He was, rather, investigating whether vanilla risk preferences a'la expected utility theory could be described as growing out of biological evolution. His finding that Nature favors agents with risk preferences that are sensitive to correlation was an unexpected consequence of this investigation—an accidentally discovered implication in an effort to model the evolutionary foundations of the already widely accepted expected utility framework. The value of evolutionary model as an engine for producing behavioral hypotheses lies precisely in its capacity to produce these accidental secondary or tertiary implications—*not* in their ability to explain what is already believed or observed. In this sense, evolution can serve as an engine of discovery for candidate behavioral hypotheses because it has the capacity to yield hypotheses that we are unlikely to posit through other means. We take our experiment's demonstration of this potential as an additional contribution of our work.

References

Andreoni, James, and John Miller. 2002. "Giving According to GARP: An Experimental Test of the Consistency of Preferences for Altruism." *Econometrica*,

[29]For this claim we cite private correspondence with Robson.

70: 737–753.

Barkow, Jerome H., Leda Cosmides, and John Tooby, ed. 1992. *The Adapted Mind: Evolutionary Psychology and the Generation of Culture.* Oxford University Press.

Bergstrom, Theodore. 2014. "On the Evolution of Hoarding, Risk-Taking, and Wealth Distribution in Nonhuman and Human Populations." *PNAS*, 111: 10860–10867.

Bergstrom, Theodore C. 1997. "Storage for Good Times and Bad: Of Rats and Men." *Mimeograph, UC Santa Barbara.*

Bolton, Gary, Axel Ockenfels, and Julia Stauf. 2015. "Social Responsibility Promotes Conservative Risk Behavior." *European Economic Review*, 74: 109–127.

Boyd, Robert, and Peter J. Richerson. 1985. *Culture and the Evolutionary Process.* University of Chicago Press.

Charness, Gary, and Matthew Rabin. 2002. "Understanding Social Preferences with Simple Tests." *Quarterly Journal of Economics*, 117: 817–869.

Cooper, David J., and John H. Kagel. 2015. "Other-Regarding Preferences: A Selective Survey of Experimental Results." In *Handbook of Experimental Economics, Volume 2.* , ed. John H. Kagel and Alvin E. Roth, 217–289.

Cooper, William S., and Robert H. Kaplan. 1982. "Adaptive "Coin-Flipping": A Decision-Theoretic Examination of Natural Selection for Random Individual Variation." *Journal of Theoretical Biology*, 94: 135–151.

Crosetto, Paolo, and Antonio Filippin. 2013. "The Bomb Risk Elicitation Task." *Journal of Risk and Uncertainty*, 47: 31–65.

Curry, Philip. 2001. "Decision Making under Uncertainty and the Evolution of Interdependent Preferences." *Journal of Economic Theory*, 98: 357–369.

Epstein, Larry, and Stephen Tanny. 1980. "Increased Generalized Correlation: A Definition and Some Economic Consequences." *Canadian Journal of Economics*, 13: 16–34.

Friedman, Daniel. 1989. "The S-Shaped Value Function as a Constrained Optimum,." *American Economic Review*, 79: 1243–1248.

Friedman, Daniel. 1991. "Evolutionary Games in Economics,." *Econometrica*, 59: 637–666.

Gould, Stephen J. 1978. "Sociobiology: The Art of Storytelling." *New Scientist*, 80: 530–533.

Heller, Yuval, and Arthur Robson. 2021. "Evolution, Heritable Risk, and Skewness Loving." *Theoretical Economics*, 16: 403–424.

Heller, Yuval, and Ilan Nehama. 2023. "Evolutionary Foundation for Heterogeneity in Risk Aversion." *Journal of Economic Theory*, 208.

Holt, Charles A., and Susan K. Laury. 2002. "Risk Aversion and Incentive Effects." *American Economic Review*, 92: 1644–1655.

Kahneman, Daniel, and Amos Tversky. 1979. "Prospect Theory: An Analysis of Decision under Risk." *Econometrica*, 47: 263–292.

Kahneman, Daniel, Jack L. Knetsch, and Richard H. Thaler. 1986. "Fairness and the Assumptions of Economics." *Journal of Business*, 59: S285–S300.

Kimbrough, Erik, Nikolaus Robalino, and Arthur Robson. 2017. "Applying "Theory of Mind": Theory and Experiments." *Games and Economic Behavior*, 106: 209–226.

Lewontin, Richard C., and Daniel Cohen. 1969. "On Population Growth in a Randomly Varying Environment." *Proceedings of the National Academy of Sciences*, 62: 1056–1060.

Loewenstein, George F., Leigh Thompson, and Max H. Bazerman. 1989. "Social Utility and Decision Making in Interpersonal Contexts." *Journal of Personality and Social Psychology*, 57: 426–441.

Mehra, R, and E.C. Prescott. 1985. "The Equity Premium: A Puzzle." *Journal of Monetary Economics*, 15: 145–161.

Mohlin, Erik. 2012. "Evolution of Theories of Mind." *Games and Economic Behavior*, 75: 299–318.

Netzer, Nick. 2009. "Evolution of Time Preferences and Attitudes toward Risk." *American Economic Review*, 99: 937–955.

Nordhöy, Frode. 1962. "Group Interaction in Decision-Making Under Risk." *Doctoral dissertation, Massachusetts Institute of Technology.*

Polman, Evan, and Kaiyang Wu. 2020. "Decision Making for Others Involving Risk A Review and Meta-analysis." *Journal of Economic Psychology*, 77: 102–184.

Rayo, Luis, and Gary S. Becker. 2007. "Evolutionary Efficiency and Happiness." *Journal of Political Economy*, 115: 302–337.

Robalino, Nikolaus, and Arthur Robson. 2016. "The Evolution of Strategic Sophistication." *American Economic Review*, 106: 1046–1072.

Robatto, Roberto, and Balazs Szentes. 2017. "On the Biological Foundation of Risk Preferences." *Journal of Economic Theory*, 172: 410–422.

Robson, Arthur. 1996. "A Biological Basis for Expected and Non-expected Utility." *Journal of Economic Theory*, 68: 397–424.

Robson, Arthur. 2001a. "The Biological Basis of Economic Behavior." *Journal of Economic Literature*, 39: 11–33.

Robson, Arthur. 2001b. "Why Would Nature Give Individuals Utility Functions?" *Journal of Political Economy*, 109: 900–914.

Robson, Arthur. 2002. "Evolution and Human Nature." *Journal of Economic Perspectives*, 16: 89–106.

Robson, Arthur, and Allen Orr. 2021. "Evolved Attitudes to Risk and the Demand for Equity." *PNAS*, 118: 1–8.

Robson, Arthur, and Larry Samuelson. 2009. "The Evolution of Time Preference with Aggregate Uncertainty." *American Economic Review*, 99: 1925–1953.

Robson, Arthur, and Larry Samuelson. 2011. “The Evolutionary Foundations of Preferences.” In *The Handbook of Social Sciences,* vol. 1, Jess Benhabib, Alberto Bisin, and Matthew Jackson (Eds.): 221–310.

Robson, Arthur, and Larry Samuelson. 2019. “Evolved Attitudes to Idiosyncratic and Aggregate Risk in Age-Structured Populations.” *Journal of Economic Theory*, 181: 44–81.

Robson, Arthur, and Larry Samuelson. 2022. “The Evolution of Risk Attitudes with Fertility Thresholds.” *Journal of Economic Theory*, 205.

Rohde, Ingrid M.T., and Kirsten I.M. Rohde. 2011. “Risk Attitudes in a Social Context.” *Journal of Risk and Uncertainty*, 43: 205–225.

Sadowski, Philipp, and Todd Sarver. 2023. “Adaptive Preferences: An Evolutionary Model of Non-Expected Utility and Ambiguity Aversion.” *Mimeograph, Duke University.*

Samuelson, Larry. 2004. “Information Based Relative Consumption Effects.” *Econometrica*, 72: 93–118.

Shupp, R.S., and Arlington W. Williams. 2008. “Risk Preference Differentials of Small Groups and Individuals.” *Economic Journal*, 118: 258–283.

Steiner, Jakub, and Colin Stewart. 2016. “Perceiving Prospects Properly.” *American Economic Review*, 106: 1601–1631.

Stoner, James A.F. 1961. “A Comparison of Individual and Group Decisions Involving Risk.” *Unpublished Manuscript.*

Stoner, James A.F. 1968. “Risky and Cautious Shifts in Group Decisions.” *Journal of Experimental Psychology*, 4: 442–459.

Tversky, Amos, and Daniel Kahneman. 1992. “Advances in Prospect Theory: Cumulative Representation of Uncertainty.” *Journal of Risk and Uncertainty*, 5: 297–323.

Online Appendices

A Theoretical Results

A.1 Proof of Propositions 1-3 from the Main Paper

A.1.1 Attitudes to Correlation

The following definitions (Definitions A.1 and A.2), and the result that follows (Theorem A.1), come from Epstein and Tanny (1980). (These are stated informally in Section 2.1 of the main paper.) In the following, capital letters denote random variables and lower case letters denote their realizations.

Definition A.1 *Consider random variables $X = (X_1, X_2)$, and $Y = (Y_1, Y_2)$ with pdfs f, and g, respectively. Say that Y_1 and Y_2 are* more correlated than *X_1 and X_2 if* (1) *X_r and Y_r have the same marginal distribution, $r = 1, 2$, and* (2)

$$P\{Y_1 \geqslant \alpha_1, Y_2 \geqslant \alpha_2\} \geqslant P\{X_1 \geqslant \alpha_1, X_2 \geqslant \alpha_2\},$$

for each positive α_1 and α_2.[30]

Definition A.2 *Consider an individual with utility $u(x_1, x_2)$. The individual is* averse to correlation *in x_1 and x_2 if*

$$E[u(X_1, X_2)] \geqslant E[u(Y_1, Y_2)], \tag{A.3}$$

whenever Y_1 and Y_2 are more correlated than X_1 and X_2. The individual has an affinity for correlation *in x_1 and x_2 if the above inequality is reversed whenever Y_1 and*

[30] Epstein and Tanny (1980) prove that condition (2) is equivalent to

$$P\{Y_1 \leqslant \alpha_1, Y_2 \leqslant \alpha_2\} \geqslant P\{X_1 \leqslant \alpha_1, X_2 \leqslant \alpha_2\},$$

for each positive α_1 and α_2. Hence, Y_1 and Y_2 are more correlated because (Y_1, Y_2) is more likely to yield pairs (x_1, x_2) where x_1 and x_2 are both high, or both low.

Y_2 are more correlated than X_1 and X_2. The individual is indifferent to correlation in x_1 and x_2 if it holds with equality whenever Y_1 and Y_2 are more(or less) correlated than X_1 and X_2.

Theorem A.1 *Consider an individual with utility $u(x_1, x_2)$, where u is twice differentiable. The individual is averse to correlation in x_1 and x_2 if and only if $u_{12} \leqslant 0$. The individual has an affinity for correlation in x_1 and x_2 if and only if $u_{12} \geqslant 0$, $r \neq s$. The individual is indifferent to correlation in x_1 and x_2 if and only if $u_{12} = 0$.*

We extend the above definitions and Theorem A.1 to utility functions with more than two arguments. Consider an individual with an interdependent utility function $u(x, \boldsymbol{x})$, where x is own consumption and $\boldsymbol{x}$ is a profile of consumption levels of some other individuals.

Definition A.3 *Consider random variables $(X, \boldsymbol{X})$, and $(Y, \boldsymbol{Y})$, where X and Y take values in $\mathscr{X} \subseteq \mathbb{R}_+$, and $\boldsymbol{X}$, $\boldsymbol{Y}$ take values in $\mathscr{X}^n$. Suppose $(X, \boldsymbol{X})$, and $(Y, \boldsymbol{Y})$, have distributions F, and G, respectively. Say that Y and $\boldsymbol{Y}$ are* more correlated *than X and $\boldsymbol{X}$ if* (1) *the marginal distributions of X and Y are the same, and similarly for $\boldsymbol{X}$ and $\boldsymbol{Y}$, and* (2) *there is some non-decreasing $h : \mathscr{X}^n \longrightarrow \mathbb{R}$ such that*

$$P\{Y \geqslant \alpha_1, h(\boldsymbol{Y}) \geqslant \alpha_2\} \geqslant P\{X \geqslant \alpha_1, h(\boldsymbol{X}) \geqslant \alpha_2\},$$

for each positive α_1 and α_2.

When $n = 1$ this corresponds to the definition of "more correlated" by Epstein and Tanny (1980), with h being the identity function on $\mathscr{X}$.

Correlation attitudes in this case then follow as in Definition A.2.

Definition A.4 *An individual is* aggregate correlation averse (ACA) *if she has interdependent utility $u(x, \boldsymbol{x})$, where*

$$E[u(X, \boldsymbol{X})] \geqslant E[u(Y, \boldsymbol{Y})], \tag{A.4}$$

whenever Y and $\boldsymbol{Y}$ are more correlated than X and $\boldsymbol{X}$. The individual is aggregate correlation loving (ACL) *if* (A.4) *is reversed whenever Y and $\boldsymbol{Y}$ are more correlated than X and $\boldsymbol{X}$. The individual is aggregate correlation neutral* (ACN) *if* (A.4) *holds with equality whenever Y and $\boldsymbol{Y}$ are more(or less) correlated than X and $\boldsymbol{X}$.*

The analysis in Epstein and Tanny (1980) straightforwardly extends to the present case, and thus we have—

Theorem A.2 *Suppose an individual has interdependent utility, $u(x, \boldsymbol{x})$, where $x \in \mathscr{X}$, and $\boldsymbol{x} \in \mathscr{X}^n$, and u is differentiable. The individual is* ACA(ACL)[ACN] *if and only if there is some non-decreasing $h : \mathscr{X}^n \longrightarrow \mathbb{R}$ such that*

$$u_x(x, \boldsymbol{x}) \geqslant (\leqslant) \, [=] \, u_x(x, \bar{\boldsymbol{x}}), \tag{A.5}$$

whenever $h(\bar{\boldsymbol{x}}) \geqslant h(\boldsymbol{x})$.

A.1.2 Proof of Proposition 1 (Aggregate Correlation Aversion)

Following Robson (1996) (as described in Section 2 of the main paper), consider a large population, where each individual chooses a gamble over consumption from a fixed set $\mathscr{F}$. Gambles imply distributions over levels of consumption, where these distributions depend on an the aggregate state of the environment. As in Curry (2001) a type of individual corresponds to how the individual conditions choice of own gamble on the distribution of choices of the others in the population. Suppose consumption of $x \in [0, \infty)$ yields j offspring with a given exogenous probability, $p(j, x)$. Assume numbers of offspring are independent across individuals, and independent of the environment, conditional on each realized profile of consumption levels.

Curry (2001) proves that the choices of the evolutionarily dominant type are consistent with expected utility maximization with interdependent utility function over offspring given by,

$$R(j, J) = \frac{j}{J}, \tag{A.6}$$

where j is own offspring and J is the number of offspring per individual in the population. Since, realized offspring are independent across individuals given each profile of realized consumption levels, it follows that choices over consumption gambles, for the dominant type, are consistent with expected utility maximization with the interdependent utility function

$$v(x, \boldsymbol{x}) = E\left[\frac{1}{J} \,\middle|\, \boldsymbol{x}\right] \cdot \left[\sum_j j \cdot p(j, x)\right], \tag{A.7}$$

where x is own consumption, and $\boldsymbol{x} \in [0, \infty)^\infty$ is the profile of consumption levels of the others in the population.

Assume—

Assumption A.2 (1) $p(j, x)$ *is continuously differentiable with respect to* x, *for each* j. (2) *The support of* $p(\cdot, x)$ *is uniformly bounded in* x. *That is,* $p(j, x) = 0$ *for all* x, *for all* $j > \bar{j}$, *for some finite* $\bar{j}$.

Next, assume that more consumption unambiguously improves fertility.

Assumption A.3 *Given any levels of consumption* x *and* $\bar{x}$, *if* $\bar{x} < x$, *then the distribution* $\{p(\cdot, x)\}$ *first-order stochastically dominates* $\{p(\cdot, \bar{x})\}$.

We then have—

Lemma A.1 *The choices of the evolutionarily dominant type are consistent with aggregate correlation aversion, i.e., aversion to correlation in* x *and* $\boldsymbol{x}$.

Proof. Recall that for the evolutionarily dominant type gamble choices are as if motivated by the interdependent utility function $v(x, \boldsymbol{x})$ (equation (A.7). Write $w(x) = \sum_j j \cdot p(j, x)$. Assumptions A.2 implies w is differentiable. We have

$$v_x(x, \boldsymbol{x}) = E\left[\frac{1}{J} \,\middle|\, \boldsymbol{x}\right] \cdot w'(x). \tag{A.8}$$

Assumption A.3 implies $w'(x) > 0$ and, moreover, that $E[1/J|\boldsymbol{x}]$ is decreasing in $\boldsymbol{x}$. The result then follows from Theorem A.2. Aversion to correlation in others' payoffs plays a role when individuals can choose their environments (see Section A.2.3 below). Evolutionarily optimal such choices must maximize the expectation of Nature's utility, which is consistent with an aversion to correlation in realized offspring for any individuals. ■

A.1.3 Proof of Proposition 2 (Aggregate Risk Aversion)

Again consider an individual with utility $u(x, \boldsymbol{x})$, where x is own consumption and $\boldsymbol{x}$ is the profile of consumption of the others. Suppose u is increasing in x. Consider distributions, F and G, over the pair $(x, \boldsymbol{x})$, where these imply the same marginal distributions, say, F_x, and $F_{\boldsymbol{x}}$, over x and $\boldsymbol{x}$, respectively. Suppose F induces less correlation in x and $\boldsymbol{x}$ than does G (in the sense of Definition A.3). If u has ACA then

$$E_F[u(x, \boldsymbol{x})] \geqslant E_G[u(x, \boldsymbol{x})]. \tag{A.9}$$

With this in mind, consider *partial certainty equivalents*, x_F, and x_G, for the gambles F, and G, respectively, where these are sure levels of own consumption that satisfy—

$$\begin{aligned} E_{F_{\boldsymbol{x}}}[u(x_F, \boldsymbol{x})] &= E_F[u(x, \boldsymbol{x})], \text{ and} \\ E_{F_{\boldsymbol{x}}}[u(x_G, \boldsymbol{x})] &= E_G[u(x, \boldsymbol{x})]. \end{aligned} \tag{A.10}$$

Here, the cdf on the left is the marginal distribution of $\boldsymbol{x}$, the profile of consumption of the others, that is implied by F. For example, the individual is indifferent between the gamble F, and receiving a sure amount, x_F, given that the others' consumption is marginally distributed in the same manner as it is under F.

In view of (A.9) and (A.10), and since u is increasing in x, we see that $x_F \geqslant x_G$ whenever the individual is ACA. Clearly, $x_F \leqslant x_G$ if the individual instead is ACL, and $x_F = x_G$ whenever the individual is ACN. We have thus proved—

Lemma A.2 *In the sense of partial certainty equivalents, an ACA(ACL) individual is more(less) averse to risk in own consumption when own consumption is more correlated with the consumption of others. For an ACN individual attitudes to risk in own consumption are independent of any correlation between x and $\boldsymbol{x}$, in the sense that such an individual has $x_F = x_G$.*

This gives Proposition 2 from the main paper.

A.1.4 Proof of Proposition 3 (IAIR) from the Main Paper

Curry (2001) already shows that for the evolutionarily dominant type preferences for idiosyncratic gambles that are independent of the distribution of others' gambles. We give the details here.

Let $\mathscr{F}_I$ be a set of idiosyncratic gambles over consumption levels. That is, for each $\pi \in \mathscr{F}_I$ the distribution of realized consumption is independent of the aggregate state of the environment.

Recall that choices over consumption gambles, for the dominant type, are consistent with expected utility maximization with the interdependent utility function

$$v(x, \boldsymbol{x}) = E\left[\frac{1}{J} \,\middle|\, \boldsymbol{x}\right] \cdot w(x),$$

where $w(x) = \sum_j j \cdot p(j, x)$, is expected offspring given consumption of x. Write $h(\boldsymbol{x}) = E[1/J|\boldsymbol{x}]$. Suppose $\boldsymbol{x}$ is distributed according to μ. Then, independence of x and $\boldsymbol{x}$ imply that our individual prefers $\pi \in \mathscr{F}_I$ to $\eta \in \mathscr{F}_I$ if

$$E_\mu[h(\boldsymbol{x})] \cdot E_\pi[w(x)] \geqslant E_\mu[h(\boldsymbol{x})] \cdot E_\eta[w(x)],$$

that is, if

$$E_\pi[w(x)] \geqslant E_\eta[w(x)].$$

Clearly the individual's preferences over $\mathscr{F}_I$ are consistent with expected utility maximization with utility $w(x)$ (expected offspring given x), independent of the distribution of choices of the others. This proves Proposition 3 from the main paper.

A.2 ARH in the Bomb Task

This section analyzes the behavioral implications of the ARH for a decision maker (DM) in our bomb tasks. To economize on notation, in the remainder, write LD for

$LEAD$, PS for POS and NG for NEG. We consider a continuous approximation of the bomb task described in the paper, where the DM can choose any box-height $b \in [0, 1]$, and bombs are uniformly distributed on $[0, 1]$. Suppose there are n individuals involved in a task (i.e., $n = 31$ in our experiments). For each n, enumerate the individuals, $i = 1, \ldots, n$, so that the DM is $i = 1$. Consider the various tasks from the main paper.

Definition A.5 ($LEAD$ Environment) *In the lead (LD) environment with n individuals the DM chooses $b_{LD}(n) \in [0, 1]$. The state s is then uniformly drawn from $[0, 1]$. If $s \leqslant b_{LD}(n)$, then each individual's payoff is $x_i = 0$, otherwise each individual's payoff is $x_i = b_{LD}(n)$.*

Definition A.6 (ID Environment) *In the independent (ID) environment with n individuals the DM chooses $b_1(n) \in [0, 1]$. Suppose the profile of choices of the others is $\boldsymbol{b}(n) = (b_2, \ldots, b_n) \in [0, 1]^{n-1}$. Each $i = 1, \ldots, n$ then draws her own state, s_i, from the uniform distribution on $[0, 1]$; these draws are statistically independent. If i has $s_i \leqslant b_i$, then her payoff is $x_i = 0$, otherwise her payoff is $x_i = b_i(n)$.*

Definition A.7 (POS Environment) *In the positive correlation (PS) environment with n individuals the DM chooses $b_1(n) \in [0, 1]$ after observing the profile of choices of the others, $\boldsymbol{b}(n) = (b_2, \ldots, b_n)$. An aggregate state s is drawn uniformly from $[0, 1]$. If i has $s \leqslant b_i$, then her payoff is $x_i = 0$, otherwise her payoff is $x_i = b_i(n)$.*

Definition A.8 (NEG Environment) *In the negative correlation (NG) environment with n individuals the DM chooses $b_1(n) \in [0, 1]$ after observing the profile of choices of the others, $\boldsymbol{b}(n) = (b_2, \ldots, b_n)$. An aggregate state s is drawn uniformly from $[0, 1]$. If the DM has $1 - s \leqslant b_1(n)$, then her payoff is $x_i = 0$, otherwise her payoff is $x_i = b_i(n)$. For the other individuals, if i has $s \leqslant b_i$, then her payoff is $x_i = 0$, otherwise her payoff is $x_i = b_i(n)$.*

A.2.1 Hypothesis 1 (Risk [Box Height] Choice in LD)

Here we consider evolutionary environments corresponding to ID and LD, in the large population limit, i.e., as n tends to infinity. As above in Section 1 suppose consumption of x results in j offspring with probability $p(j,x)$. As above assume realized numbers of offspring are independent given each profile of consumption levels.

Notice that expected offspring of an individual that consumes x, is $w(x) = \sum_j j \cdot p(j,x)$. In ID expected offspring of an individual with a box height of $b \in [0,1]$ is then

$$W(b) = (1-b) \cdot w(b) + b \cdot w(0). \tag{A.11}$$

In the large population limit the evolutionary dominant type chooses b_{ID} where b_{ID} maximizes $W(b)$. In the remainder assume that b_{ID} is a unique critical point of W.

Assumption A.4 *$W(b)$ has a unique local maximum in* $[0,1]$.[31]

In the evolutionary environment corresponding to LD, one individual (the leader) makes a box choice for the entire large population. Consider a type of leader that chooses b in each period. In the LD environment if the aggregate state is $s \leqslant b$, then each individual in the population consumes 0. Since the population is large, offspring per individual in the population in this case is $w(0)$. On the other hand if $s > b$, and the population is large, then offspring per individual is $w(b)$. It follows that the growth rate of the population where the leader in each period chooses b is

$$\Phi(b) = (1-b) \cdot \log(w(b)) + b \cdot \log(w(0)). \tag{A.12}$$

(See the discussion leading to the characterization of the long run-growth rate when there is aggregate risk in Section 2 of the main paper.) Suppose b_{LD} is the maximizer of $\Phi(b)$.

Recall that $w(x)$ is continuous on $[0,1]$ (Assumptions A.2), and thus $W(0) = W(1) =$

[31] For example, a sufficient, though not necessary, condition for this to hold is that $w(x)$ is log-concave. To see that log-concavity is not necessary, notice that $w(x) = \exp\{x^2\}$ is not log-concave, but W in this case has a unique critical point in $(0,1)$ which is the global maximizer of W in $[0,1]$.

0. Moreover, $w(x)$ is non-negative and increasing on $[0, 1]$ and therefore $W(x) > 0$ for some $x \in [0, 1]$. It follows that b_{ID} must be an interior point of $[0, 1]$. Continuous differentiability of W, and Assumption A.4, imply that b_{ID} is the unique critical point of W in $[0, 1]$ and, moreover, that W is increasing on $[0, b_{ID})$ and decreasing on $(b_{ID}, 1]$. With this in mind, notice that $\log(w(x))$ is more concave than $w(x)$ in the sense that $\log(w)$ has a higher Arrow-Pratt coefficient of aversion to risk in x than does w, for each $x \in [0, 1]$. That $b_{LD} < b_{ID}$ is then implied by the following basic result. The proof is given later in Section A.3

Lemma A.3 *Consider*

$$U(b) = (1 - b) \cdot u(b) + b \cdot u(0),$$

and

$$V(b) = (1 - b) \cdot v(b) + b \cdot v(0),$$

where $u : [0, 1] \longrightarrow \mathbb{R}$ *and* $v : [0, 1] \longrightarrow \mathbb{R}$ *are increasing, and continuously differentiable on* $[0, 1]$. *Suppose* u *is more concave than* v. *Then,* $U'(b) \geqslant 0$ *on* $S \subseteq [0, 1]$ *implies* V *is increasing on* S.

A.2.2 Hypotheses 2-4 (Choices in PS and NG)

Consider a DM in the ID, PS, and NG tasks (as described above in Definitions A.6-A.8). Fix the number of individuals involved in the tasks, and suppose the DM's choice is consistent with expected utility maximization for some interdependent utility function, $u(x, \boldsymbol{x})$, where x is the payoff of the DM and $\boldsymbol{x}$ is the profile of payoffs of the others. Assume—

Assumption A.5 $u(x, \boldsymbol{x})$ *is non-negative, and, moreover, increasing and continuously differentiable in* x, *for each* $\boldsymbol{x}$.

Our hypotheses (H2-H4) regarding PS and NG choices, and how these are related to ID choice, from Section 4 of the main paper, are derived from the following key result.

Proposition A.4 *Each of the following is true—*

(1) *If the DM is ACN, then $b_{PS} = b_{ID} = b_{NG}$.*

(2) *If the DM is ACL and has IAIR, then $b_{NG} \leqslant b_{ID} \leqslant b_{PS}$.*

(3) *If the DM is ACA and has IAIR, then $b_{PS} \leqslant b_{ID} \leqslant b_{NG}$.*

(4) *If the DM is ACL, and $b_{NG} \leqslant 1/2$, then $b_{NG} \leqslant b_{PS}$.*

(5) *If the DM is ACA, and $b_{PS} \leqslant 1/2$, then $b_{NG} \geqslant b_{PS}$.*

Proof. In PS and NG there is an aggregate state, s, which is uniformly distributed on $[0,1]$. For each s, the payoffs of the others in PS and NG is $\boldsymbol{x}(s) = (x_2(s), \ldots, x_n(s))$, where $x_i(s) = 0$ if $b_i \leqslant s$, and $x_i(s) = b_i$ if $b_i > s$, $i = 2, \ldots, n$. Recall that in PS if the DM chooses b, then her payoff is b if $b < s$, and zero otherwise. In NG her payoff is b if $b < 1 - s$, and zero otherwise. Expected utility for the DM from choosing $b \in [0,1]$ in PS is then

$$V_{PS}(b) = (1-b) \cdot E\left[u(b, \boldsymbol{x}(s)) \mid s > b\right] + b \cdot E\left[u(0, \boldsymbol{x}(s)) \mid s \leqslant b\right]. \tag{A.13}$$

In NG this is

$$V_{NG}(b) = (1-b) \cdot E\left[u(b, \boldsymbol{x}(s)) \mid s < 1-b\right] + b \cdot E\left[u(0, \boldsymbol{x}(s)) \mid s \geqslant 1-b\right]. \tag{A.14}$$

In ID it is

$$V_{ID}(b) = (1-b) \cdot E_\mu[u(b, \boldsymbol{x})] + b \cdot E_\mu[u(b, \boldsymbol{x})], \tag{A.15}$$

where μ are the subjective beliefs of the DM about the payoffs of others. These are arbitrary, subject only to the condition that under μ the payoff of the DM is independent of $\boldsymbol{x}$.

Assumption A.5 implies (1) $V_r(b)$ is continuous, $r \in \{ID, PS, NG\}$, (2) V_{ID} has a continuous derivative, and (3) V_r is has a left continuous derivative, $r \in \{PS, NG\}$, with any discontinuities occurring at the points, $b = b_i$, $i = 2, \ldots, n$. Hence, it follows that $V_r(b) = V_r(0) + \int_0^b V_r'(z)\, dz$, $r \in \{ID, PS, NG\}$. These derivatives, where continuous, are

$$\begin{aligned}\frac{dV_{PS}(b)}{db} &= (1-b)\cdot E\left[u_x(b,\boldsymbol{x}(s))\,|\,s>b\right] - \left[u(b,\boldsymbol{x}(b)) - u(0,\boldsymbol{x}(b))\right],\\ \frac{dV_{NG}(b)}{db} &= (1-b)\cdot E\left[u_x(b,\boldsymbol{x}(s))\,|\,s<1-b\right] - \left[u(b,\boldsymbol{x}(1-b)) - u(0,\boldsymbol{x}(1-b))\right],\ \text{and}\\ \frac{dV_{ID}(b)}{db} &= (1-b)\cdot E_\mu\left[u_x(b,\boldsymbol{x})\right] - E_\mu\left[u(b,\boldsymbol{x}) - u(0,\boldsymbol{x})\right].\end{aligned} \tag{A.16}$$

Notice that $V_r(0) = V_r(1) = E[u(0,\boldsymbol{x}(s))]$, $r \in \{PS, NG\}$, and $V_{ID}(0) = V_{ID}(1) = E[u(0,\boldsymbol{x})]$. Since $u(b,\boldsymbol{x})$ is non-negative and increasing in b, for all $\boldsymbol{x}$, it follows that any maximizer of V_r is an interior point of $[0,1]$, for each $r \in \{ID, PS, NG\}$.

Proof of (1). Suppose the DM is ACN. Theorem A.2 implies $u_x(x,\boldsymbol{x})$ does not depend on $\boldsymbol{x}$, and thus neither does $u(x,\boldsymbol{x}) - u(0,\boldsymbol{x})$. Since $u_x(x,\boldsymbol{x})$ is continuous for each x and $\boldsymbol{x}$ (Assumption A.5) it follows that V_r has a continuous derivative, $r \in \{ID, PS, NG\}$, and moreover, that $V'_{ID}(b) = V'_{PS}(b) = V'_{NG}(b)$, for all b. Since each V_r has an interior maximizer, it follows that any maximizer of V_r is also a maximizer of V_t, $r,t \in \{ID, PS, NG\}$.

Proof of (2). Suppose the DM is ACL and has IAIR. $\boldsymbol{x}(s)$ is non-decreasing in s (i.e., if $s \geqslant \hat{s}$, then $x_i(s) \geqslant x_i(\hat{s}), i = 2,\ldots,n$). Hence, Theorem A.2 implies,

$$\begin{aligned} E[u_x(b,\boldsymbol{x}(s))|s>b] &\geqslant u_x(b,\boldsymbol{x}(b)), \qquad \text{and}\\ E[u_x(b,\boldsymbol{x}(s))|s<1-b] &\leqslant u_x(b,\boldsymbol{x}(1-b)).\end{aligned} \tag{A.17}$$

Thus,

$$\begin{aligned}\frac{dV_{PS}(b)}{db} &\geqslant (1-b)\cdot u_x(b,\boldsymbol{x}(b)) - \left[u(b,\boldsymbol{x}(b)) - u(0,\boldsymbol{x}(b))\right],\ \text{and}\\ \frac{dV_{NG}(b)}{db} &\leqslant (1-b)\cdot u_x(b,\boldsymbol{x}(1-b)) - \left[u(b,\boldsymbol{x}(1-b)) - u(0,\boldsymbol{x}(1-b))\right].\end{aligned} \tag{A.18}$$

Since the DM has IAIR it follows that the right-hand side expressions in (A.18) are both equal to $dV_{ID}(b)/db$. This follows from Lemma A.3 because IAIR implies that for $u(x,\boldsymbol{x})$ the coefficient of aversion to risk in x does not depend on $\boldsymbol{x}$. Equations (A.18) are then—

$$\begin{aligned} \frac{dV_{PS}(b)}{db} &\geqslant \frac{dV(b, \boldsymbol{x}(b))}{db}, \quad \text{and} \\ \frac{dV_{NG}(b)}{db} &\leqslant \frac{dV(b, \boldsymbol{x}(1-b))}{db}. \end{aligned} \tag{A.19}$$

Suppose b_{ID} is a maximizer of V_{ID}. Then, for each $b, \hat{b} \in [0,1]$ such that $b < b_{ID} < \hat{b}$, it follows that

$$\begin{aligned} &\int_{b}^{b_{ID}} V'_{ID}(z)\, dz \geqslant 0. \text{ and} \\ &\int_{b_{ID}}^{\hat{b}} V'_{ID}(z)\, dz \leqslant 0. \end{aligned} \tag{A.20}$$

In view of equations (A.19) we have

$$\begin{aligned} &\int_{b}^{b_{ID}} V'_{PS}(z)\, dz \geqslant 0. \text{ and} \\ &\int_{b_{ID}}^{\hat{b}} V'_{NG}(z)\, dz \leqslant 0, \end{aligned} \tag{A.21}$$

for all $b, \hat{b} \in [0,1]$ with $b < b_{ID} < \hat{b}$. If follows that V_{PS} must have a maximizer $b_{PS} \geqslant b_{ID}$, and that V_{NG} must have a maximizer $b_{NG} \leqslant b_{ID}$.

Proof of (3). Suppose the DM is ACA and has IAIR. Theorem A.2 implies reversal of the inequalities in (A.17)-(A.19) in the above proof for an ACL DM. This gives, for the ACA case,

$$\begin{aligned} &\int_{b_{ID}}^{b} V'_{PS}(z)\, dz \leqslant 0. \text{ and} \\ &\int_{\hat{b}}^{b_{ID}} V'_{NG}(z)\, dz \geqslant 0, \end{aligned} \tag{A.22}$$

for all $b, \hat{b} \in [0,1]$ with $b < b_{ID} < \hat{b}$. If follows that V_{PS} must have a maximizer $b_{PS} \leqslant b_{ID}$, and that V_{NG} must have a maximizer $b_{NG} \geqslant b_{ID}$.

Proof of (4)-(5). Suppose the DM is ACL. Theorem A.2 implies

$$E\left[u_x(b, \boldsymbol{x}(s)) \,|\, s > b\right] \geqslant E\left[u_x(b, \boldsymbol{x}(s)) \,|\, s < 1-b\right],$$

for all $b \in [0,1]$. It also implies

$$u(b, \boldsymbol{x}(s)) - u(0, \boldsymbol{x}(s))$$

is non-decreasing in s, for each $b \in [0,1]$. In view of equation (A.16) it follows that $V'_{PS}(b) \geqslant V'_{NG}(b)$ on $[0, 1/2]$. Therefore, if $b_{NG} \leqslant 1/2$ is a maximizer of V_{NG}, then

$$0 \leqslant \int_b^{b_{NG}} V'_{NG}(z)\, dz \leqslant \int_b^{b_{NG}} V'_{PS}(z)\, dz,$$

for each $b < b_{NG}$. It follows that V_{PS} has a maximizer $b_{PS} \geqslant b_{NG}$.

Suppose instead the DM is ACA. Theorem A.2 implies

$$E\left[u_x(b, \boldsymbol{x}(s)) \,|\, s > b\right] \leqslant E\left[u_x(b, \boldsymbol{x}(s)) \,|\, s < 1 - b\right],$$

for all $b \in [0,1]$. It also implies

$$u(b, \boldsymbol{x}(s)) - u(0, \boldsymbol{x}(s))$$

is non-increasing in s, for each $b \in [0,1]$. We have $V'_{PS}(b) \leqslant V'_{NG}(b)$ on $[0, 1/2]$. Therefore, if $b_{PS} \leqslant 1/2$ is a maximizer of V_{PS}, then

$$0 \leqslant \int_b^{b_{PS}} V'_{PS}(z)\, dz \leqslant \int_b^{b_{PS}} V'_{NG}(z)\, dz,$$

for each $b < b_{PS}$. It follows that V_{NG} has a maximizer $b_{NG} \geqslant b_{PS}$. ■

A.2.3 Hypotheses 5-7 (Choice of Task)

Recall that each of our DMs participates in an ID task and also a second task, $r \in \{LD, PS, NG\}$. In a Choice Task, the DM chooses between the two experienced environments. We derive Hypotheses 5-7 by showing that the population growth rate in evolutionary environments corresponding to ID, LD, PS, and NG, is highest in the ID environment, when the population is large. With this in mind, consider evolutionary environments corresponding to $r \in \{ID, LD, PS, NG\}$, where payoffs for individuals in the population are determined as described in the corresponding Definitions A.5-A.8. As in Section A.1.2 above, suppose a payoff of x yields j offspring with probability $p(j,x)$. Again realized numbers of offspring are independent given each profile of payoffs.

Assume that in the ID environment box choices are evolutionarily optimal for this environment. When the population is large, independence of payoffs in ID implies the evolutionarily dominant type in the ID environment chooses b_{ID}, where b_{ID} maximizes expected offspring in the bomb task—

$$W(b) = (1-b)\cdot w(b) + b\cdot w(0).$$

Again, $w(x) = \sum_j j\cdot p(j,x)$ is expected offspring for an individual with a payoff of x.

Now suppose $W_r(s)$ is the number of offspring per individual in environment $r \in \{LD, PS, NG\}$ when the aggregate state of the environment is $s \in [0,1]$. Expected number of offspring per individual in environment $r \in \{LD, PS, NG\}$ is then

$$\bar{W}_r = \int_0^1 W_r(s)\,ds. \tag{A.23}$$

The growth rate in environment ID is

$$\Phi_{ID} = \log(W(b_{ID})), \tag{A.24}$$

while in $r \in \{LD, PS, NG\}$ it is

$$\Phi_r = \int_0^1 \log(W_r(s))\,ds. \tag{A.25}$$

The strict concavity of log implies

$$\Phi_r < \log(\bar{W}_r), \tag{A.26}$$

$r \in \{LD, PS, NG\}$. Clearly, $W(b_{ID}) \geqslant W_r$, for each $r \in LD, PS, NG$, and therefore $\Phi_{ID} > \Phi_r$, $r \in \{LD, PS, NG\}$.

A.3 Proof of Lemma A.3

Since u is more risk averse than v, it follows that $u(b) - u(0)$ is also more risk averse than $v(b) - v(0)$. Therefore, there is an increasing, strictly concave function ψ, with $\psi(0) = 0$, such that

$$u(b) - u(0) = \psi(v(b) - v(0)).$$

Hence—

$$U(b) = (1 - b) \cdot [\psi(v(b) - v(0)) + u(0)] + b \cdot u(0).$$

Since u, and ψ are differentiable, so is U and, moreover—

$$U'(b) = (1 - b) \cdot \psi'(v(b) - v(0)) \cdot v'(b) - \psi(v(b) - v(0)),$$

that is,

$$U'(b) \leqslant 0 \Longleftrightarrow (1 - b) \cdot v'(b) \cdot \psi'(v(b) - v(0)) - \psi(v(b) - v(0)) \leqslant 0.$$

Multiplying each side of the right-hand inequality by $q(b) = \frac{v(b)-v(0)}{\psi(v(b)-v(0))}$ gives,

$$U'(b) \leqslant 0 \Longleftrightarrow (1 - b) \cdot v'(b) \cdot \psi'(v(b) - v(0)) \cdot q(b) - [v(b) - v(0)] \leqslant 0. \tag{A.27}$$

Since ψ is strictly concave, and since $\psi(0) = 0$,

$$\psi'(v(b) - v(0)) \cdot q(b) = \frac{\psi'(v(b) - v(0))}{\frac{\psi(v(b)-v(0))}{v(b)-v(0)}} < 1. \tag{A.28}$$

By assumption, $v'(b) > 0$. In view of equations (A.27) and (A.28) we thus have

$$U'(b) < 0, \text{ if } (1 - b) \cdot v'(b) - [v(b) - v(0)] \leqslant 0.$$

This establishes the result since $V'(b) = (1 - b) \cdot v'(b) - [v(b) - v(0)]$.

B Additional Empirical Results

	LEAD	LEAD	LEAD	POS	POS	POS	NEG	NEG	NEG
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Constant	0.783***	0.777***	0.806***	0.756***	0.728***	0.732***	0.682***	0.657***	0.710***
	(0.055)	(0.069)	(0.094)	(0.059)	(0.059)	(0.060)	(0.039)	(0.049)	(0.056)
Aggregate	−0.118**	−0.117**	−0.163**	−0.127***	−0.132***	−0.133***	0.035	0.038	0.006
	(0.051)	(0.053)	(0.070)	(0.048)	(0.048)	(0.047)	(0.054)	(0.055)	(0.059)
Second		0.012	−0.052		0.061	0.028		0.046	0.018
		(0.053)	(0.070)		(0.048)	(0.047)		(0.055)	(0.059)
Observations	192	192	138	166	166	98	206	206	148
R^2	0.015	0.016	0.026	0.018	0.022	0.028	0.001	0.004	0.001
Adjusted R^2	0.010	0.005	0.011	0.012	0.010	0.008	-0.004	-0.006	-0.013
Note:									*p<0.1; **p<0.05; ***p<0.01

Table 2: Estimated drivers of CRRA parameters. *Notes: Each column reports a separate specification run on data from the treatment (LEAD, POS and NEG) listed at the column's top. Constant measures mean behavior in the ID Task. Aggregate is a dummy variable for the aggregate task (LEAD, POS or NEG depending on treatment). Second is a dummy that takes a value of 1 if the Task was experienced by the subjects second in order. Specifications (1) and (2) make use of the entire dataset, while (3) makes use of data only from the (majority of) subjects who expressed a preference for the ID task over the aggregate task at the end of the experiment.*

C Instructions to Subjects

C.1 General Instructions

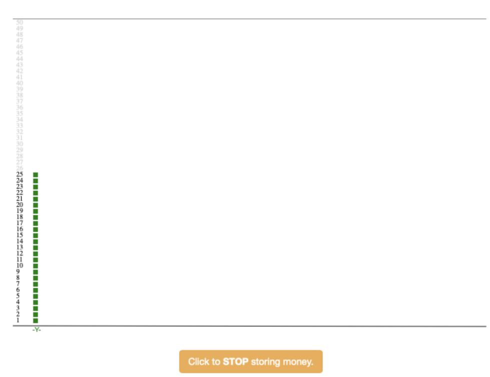
All subjects, regardless of treatment, begin with the following instructions.

Instructions

- We will start by providing you with **INSTRUCTIONS** for the study.
- We will ask you **COMPREHENSION QUESTIONS** to check that you understand the instructions. You should be able to answer all of these questions correctly.
- Please read and follow the instructions closely and carefully.
- If you **COMPLETE** the main parts of the study, you will receive a **GUARANTEED PAYMENT** of **$3.00**.
- In addition, your **CHOICES** in the **DECISIONS** portion of the study will result in **PERFORMANCE-BASED EARNINGS**. You will make several **DECISIONS** worth **REAL MONEY**. Your points from **ONE OF THE DECISIONS** will be randomly selected by computer and will be converted into an additional payment.
- Please **DO NOT REFRESH OR CLOSE YOUR BROWSER** during the study. Doing so may prevent you from completing the study.

Storing Money in a Cave

- In this experiment you will decide how much **MONEY** to store in a cave. Each unit of money is shown as a dark green box in a stack on your screen.
- **Example:** *In the example below, there are 25 dark green boxes, meaning you have stored 25 units of money.*



- The more boxes you stack in the cave, the higher the **BONUS** we will pay you at the end of the experiment -- but only if your stack of boxes isn't **DESTROYED**!

Comprehension Questions

- **How many boxes do you have stacked in the cave in the example below?**

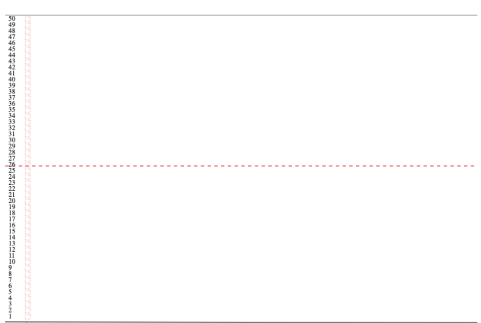
Click to **STOP** storing money.

- 10
- 18
- 25
- 35
- 45

Submit Quiz

The Random Laser

- After you have decided how many boxes to stack, a LASER BEAM will shoot from the left side of the screen (shown as a dashed red line).
- If the laser beam hits your stack of boxes (anywhere!), your **ENTIRE** stack of money will be **DESTROYED** (it will turn hollow and red) and you will earn **NOTHING ($0).**
- If your stack is **NOT** hit by the laser, you will earn **a HIGHER BONUS** the more boxes you've stacked up.
- **Example:** *In the example below, you decided to store 50 boxes. The laser struck from 25 boxes up and therefore hit your stack, so you earn nothing.*



Your money is destroyed and you earn **$0.**

- You will **NOT KNOW** the vertical location of the laser while you decide how much money to store in the cave -- it could fire from any height from 0 to 50 boxes up, and will strike at each possible height with equal likelihood. This means, the higher your stack is (i) the **MORE** you might earn but also (ii) the **HIGHER** the likelihood it will be struck by the laser and you will earn nothing.

Comprehension Questions

Click to **STOP** storing money.

- **There are 18 boxes in the example above. How many units of money will you earn if the laser fires from 10 boxes up?**

- 0
- 10
- 18
- 25

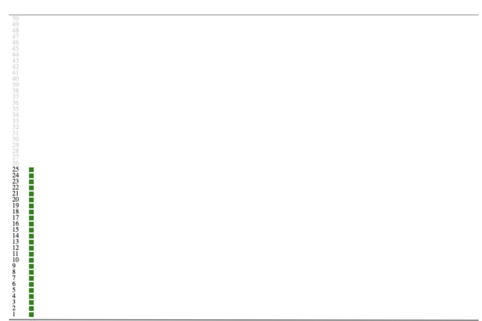
- **There are 18 boxes in the example above. How many units of money will you earn if the laser fires from 25 boxes up?**

- 0
- 10
- 18
- 25

Submit Quiz

Making Your Decision

- To start stacking boxes in the cave, press the **BUTTON AT THE BOTTOM**. Boxes will immediately and automatically appear in the cave, one at a time. Then, press the button again **TO STOP** stacking boxes.
- A **COUNTER** (a series of numbers) on the left side of the screen will count up to 50 (even after you've stopped stacking boxes), by turning the gray numbers black, one at a time. Once the counter hits 50, the Task is over, and the laser will fire, determining whether you earned money.



Click to **STOP** storing money.

- We will have you play **SEVERAL** of these **TASKS**. In each task, the laser's position will be randomly determined by the computer. That means, you will never know in any Task where the laser will show up!
- At the end of the experiment we will **RANDOMLY SELECT ONE** Task (across all of the Tasks you play) to pay you based on. This means, you should treat each individual Task as if it is the **ONLY ONE** that will count.
- We will pay you a bonus of **$0.10** for every unit of money you earn (e.g. every box stacked in the cave that is not destroyed) in the randomly selected Task.

Comprehension Questions

- **In each Task, the laser's vertical position will be**

○ the same in every Task
○ different in each Task but will follow a predictable pattern
○ random in each Task and will therefore be unpredictable

- **Suppose you've stacked 20 boxes in the cave and it wasn't destroyed by the laser. What bonus will we pay you if that Task is selected for payment.**

○ $20.00
○ $10.00
○ $0.20
○ $0.10
○ $2.00
○ $1.00

Submit Quiz

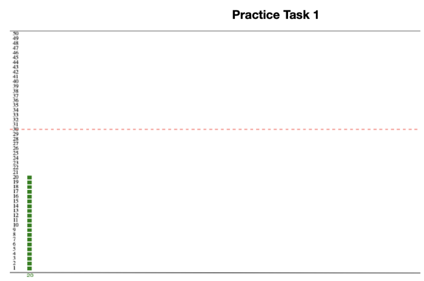
C.2 ID Task

If the ID task is randomized to be first-assigned, the subject immediately plays two practice periods (for no payment) of the ID treatment. They receive these instructions:

Practice Tasks

- We are now going to give you two **PRACTICE TASKS**, so you can get a better feel for the Task you are about to face.
- In each of these practice tasks you will stack boxes in the cave, as just described in the instructions.
- However, when you are done stacking your boxes, you will press a red button **several times** to see several, hypothetical laser blasts.
- In the actual experiment these laser blasts will be **RANDOMLY** determined by the computer (each height from 1-50 will be equally likely). But in these practice periods they will **NOT** be random but will instead be selected to give you a sense of how different laser heights could effect your earnings differently.
- Your decisions in these practice tasks **WILL NOT IMPACT** your actual earnings. The purpose of these practice tasks is just to give you a sense of the effects of the laser's position and how it can impact your earnings.

In each of the two practice periods, the subject makes a choice as normal, and then observes the effect of four draws of s on payments. Here is an example screenshot.



If the laser is 30:

Your money survives and you earn **20** units of money.

Try again, to see what another hypothetical laser would do:

See a Random, **HYPOTHETICAL** Laser

The instructions close as follows before reaching the subject's actual ID task.

Ready to Play for Real?

- We are now going to give you the **real Task**.
- You will stack boxes in the cave just as in the Practice Tasks.
- However, as soon as you are finished, the laser will randomly fire **once** and (if this Task is selected for payment) will determine your **actual bonus payment in the experiment**.
- Unlike in the practice periods, the laser's position will be **RANDOMLY** determined by the computer (and can take any value from 1-50 with equal likelihood).

C.3 Aggregate Risk Treatment

After the ID treatment (if ID is assigned first) or before it (otherwise), the subject is given instructions for one of the three aggregate risk treatments (LEAD, POS or NEG) and is given two practice periods that are exactly the same as the one's for ID but under the aggregate risk treatment they are assigned.

If the subject is assigned to the LEAD treatment, she receives the following instructions

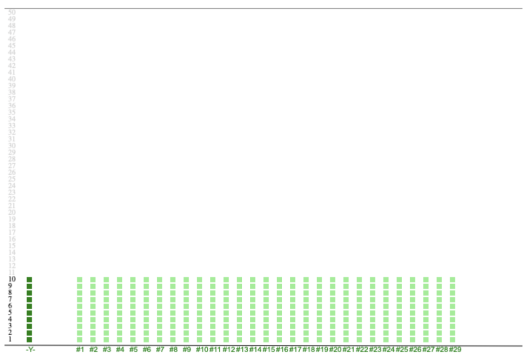
Choosing for Everyone

- In the next set of several Tasks, you will share the cave with **THIRTY OTHER PEOPLE**. These other people are **ACTUAL PARTICIPANTS** in this experiment who will also have their own stacks of money in the cave that will determine their earnings.

- We will show you the other people's stacks as LIGHTER GREEN STACKS of boxes: **EACH STACK** shows the choice made for a **DIFFERENT PARTICIPANT** who is sharing your cave.

- Your choice will determine not only your own stack, but **other people's stack as well**. When you press the button, your stack will start growing but so will everyone else's stack. When you press the button again, all of your stacks will stop growing.

- **Example:** *In the example below, you and each other person in the cave have stacked 10 units of money so far. When you click the button at the bottom, all of you will stop stacking money and will have exactly the same sized stacks.*



Click to **STOP** storing money.

- Afterwards, you and the other participants will also face exactly the SAME LASER, which will fire from exactly the same vertical position for all of you. If any person's (including your) stack is struck by the laser, you will **ALL EARN $0** instead of the amount you have each stacked. Otherwise, you will each earn $0.10 for every box you have in your own stack.

- Each of the thirty other participants will also be making a choice of how many boxes to stack for the group. When the experiment is finished, we will **RANDOMLY** select one of your choices to determine the **ENTIRE GROUP'S** payment. That means you should treat this Task as though your choice is determining the payments of every person in the group -- after all, it might be!

Comprehension Questions

- **You and the other 30 participants will**

○ Earn the same amount of money.

○ Each earn a different amount of money.

Submit Quiz

If the subject is assigned to the POS or NEG treatment, she receives the following instructions:

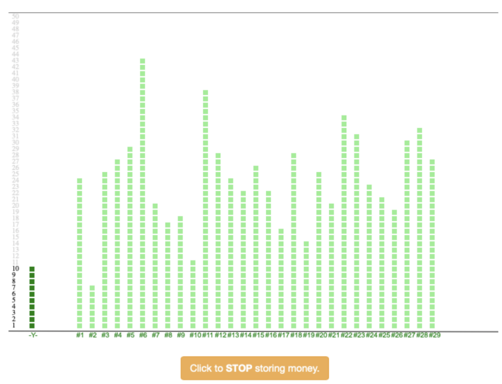
Sharing the Cave With Other People

- In the next set of several Tasks, you will share the cave with **THIRTY OTHER PEOPLE**. These are **ACTUAL PARTICIPANTS** in this experiment who have **ALREADY DECIDED** how much money to stack in the cave and who are being paid in exactly the same way you are.

- We will show you the other people's stacks as LIGHTER GREEN STACKS of boxes: **EACH STACK** shows the choice made for a **DIFFERENT PARTICIPANT** who is sharing your cave.

- Your choices will have no effect on other people's earnings, and their choices will have no effect on yours. You (and the other participants) will be paid purely based on the number of boxes you stack.

- **Example:** *In the example below, you have stacked 10 units of money so far (shown in dark green above the letter 'Y,' which stands for 'you'). The other players (player #1, #2 etc.) and their choices are shown as light green stacks.*



- You and the other participants will face exactly the SAME LASER, which will fire from exactly the same vertical position for all of you. If any person's (including your) stack is struck by the laser, they **EARN $0** for the Task instead of the amount they have stacked.

- When each Task begins, you will see a group of other people's stack decisions on your screen. You then press the button at the bottom to determine YOUR OWN STACK in the cave.

Comprehension Questions

- **Your decision**

○ will affect the 30 other participants' earnings.
○ will affect only your own earnings.
○ will affect both your own earnings and other participants' earnings.

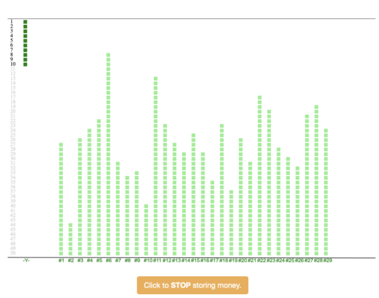
Submit Quiz

If the subject is assigned the NEG treatment she receives

From the Ceiling

- In the next several Tasks, **YOUR STACK** will be **HUNG FROM THE CEILING** of the cave instead of being stacked on the floor. That is, other participants will store their money on the floor, but you will hang yours from the ceiling.

- **Example:** *In the example below, you decided to store 10 boxes. But your boxes (shown in DARK GREEN) are "stacked" from the top of the cave, rather than from the bottom of the cave. Notice that the other participants's boxes are still stacked from the floor.*



- Nonetheless, you and the other participants are paid in the same way. The laser will shoot from some random vertical position and if it strikes your or anyone else's stack, that person will earn nothing.

Comprehension Questions

- **In the next few Tasks**

○ Everyone's money will be hung from the ceiling.
○ Your money will be stacked on the floor and everyone else's will be hung from the ceiling.
○ Your money will be hung from the ceiling and everyone else's will be stacked on the floor.
○ Everyone's money will be stacked on the floor.

Submit Quiz

C.4 Choice Task

Finally, at the end of the experiment, the subject is given a Choice task corresponding to the treatment. In the LEAD treatment she receives the following task:

Choose Your Final Task

For the final set of Tasks, we will allow you to **choose which version** of the task you will face. Click on the version of the task **you would prefer** to face in the final set of tasks.

- Make the choice in a cave on your own
- Make the choice both for yourself and for 30 other people.

Note: If you choose **not** to make a choice for 30 other participants, this will **not** reduce these participants' opportunity to earn money in the experiment. It just means their earnings will not be determined by your choices in the Task.

In the POS treatment she receives the following task:

Choose Your Final Task

For the final set of Tasks, we will allow you to **choose which version** of the task you will face. Click on the version of the task **you would prefer** to face in the final set of tasks.

- Make the choice in a cave on your own, facing your own random laser
- Make the choice in a cave with 30 other people, facing the same laser these other people face.

In the NEG treatment she receives the following task:

Choose Your Final Task

For the final set of Tasks, we will allow you to **choose which version** of the task you will face. Click on the version of the task **you would prefer** to face in the final set of tasks.

- Make the choice in a cave on your own, facing your own random laser
- Make the choice in a cave with 30 other people, facing the same laser these other people face. Your boxes will once again be hung from the ceiling, just as before.