LAURA J. KORNISH and KARL T. ULRICH*

How important is the original conception of an idea—the "raw" idea—to an innovation's success? In this article, the authors explore whether raw ideas judged as "better" fare better in the market and also determine the strength of that relationship. The empirical context is Quirky.com, a community-driven product development company for household consumer products. The data include descriptions of the raw ideas as originally proposed, the ultimate product designs that resulted from those ideas, and sales figures. In addition, they contain two measures of idea quality: those from online consumer panelists and those from expert evaluators. The authors note the following findings: First, online consumer panels are a better way to determine a "good" idea than are ratings by experts. Second, predictions with samples as small as 20 consumers are reliable. Third, there is a stronger predictive link between raw ideas and consumers' purchase intent of final product designs than there is between those intentions and market outcomes. Fourth, the commercial importance of the raw idea is large, with ideas one standard deviation better translating to an approximately 50% increase in sales rate.

Keywords: innovation, new product development, raw ideas, market outcomes, new product evaluations

Online Supplement: http://dx.doi.org/10.1509/jmr.12.0401

The Importance of the Raw Idea in Innovation: Testing the Sow's Ear Hypothesis

A "French fry restaurant" in a college town recently closed its doors. The phrase "French fry restaurant" captures the idea underlying the eatery: it was a fast-casual dining establishment that offered meals incorporating French fries—fries with sausage, fries with Thai toppings, fries

with chili, and so on. This restaurant's failure may not be surprising. Perhaps readers are asking, "Is that restaurant concept even a good idea?"

This article explores how important the quality of the raw idea is in determining success in innovation. On the one hand, it could be argued that without a good idea, a product's chance of success is very small. The eighteenth-century author Jonathan Swift reportedly coined the adage "You can't make a silk purse out of a sow's ear." The "sow's ear hypothesis" is that there is a demonstrable connection between the quality of the raw idea and the success of the resultant innovation. On the other hand, some could argue that with the right resources and approach, an innovator can create value out of just about anything. (Indeed, in 1921, engineers at consulting firm Arthur D. Little literally spun a silk purse out of processed sows' ears [Arthur D. Little Inc. 1921].) This view might be called the "Midas hypothesis," or the perspective that ideas are overrated and execution is

what matters (Ries 2011). Bolstering the Midas hypothesis is the recognition that raw ideas are rarely truly novel (Johnson 2010; Kornish and Ulrich 2011): if the raw idea is not unique, how can it be the primary source of value?

We define an "idea" as an opportunity to create value through further investment (Terwiesch and Ulrich 2009). Ideas can take different forms in innovation. An idea may be the recognition of a new need; for example, in the early 1990s, search engine innovators recognized the need to navigate the vast amount of information emerging on the worldwide web. Or an idea may be a new concept providing a solution to an existing need; the idea underlying the Snuggie was a blanket with sleeves for the user's arms, thus creating a new way to stay warm. An idea may also be the conjecture that an existing solution could meet an emerging need; for example, the idea that sparked the Macintosh computer was that the graphical user interface pioneered by Xerox as a corporate word-processing tool could address the public's burgeoning desire to access the power of computing.

Ideas evolve over the course of the innovation process. We define the "raw idea" as the opportunity as conceived at the outset of the innovation effort in a specific organizational context. In most cases, the raw idea as it first enters an organization's innovation process existed in an even more rudimentary state in the mind of the originator. A raw

idea for an innovation is often expressed in words or with a simple visual depiction. For example, Figure 1 contains a visual depiction of a raw idea from our data set along with the corresponding final design that was developed from the raw idea. In our empirical setting, the raw ideas were the actual ideas proposed at the beginning of a structured innovation process. For other contexts, the raw idea could take a different form and the description might be more or less elaborate. In a pharmaceutical innovation process, the raw idea might be a newly synthesized compound, described fully by its molecular structure. In a movie studio, the raw idea might be a one-sentence plot description.

Our conceptual framework recognizes that an innovation's value can originate from several sources: (1) the idea itself, (2) decisions made during the development and marketing of the idea, and (3) exogenous factors. We examine the roles of these sources in creating value. With this framework, we can answer our central research question: To what extent does the raw idea determine innovation success? We answer this question in two ways. First, we analyze how much of the variation in innovation outcomes is explained by variation in the quality of the raw ideas. Second, we investigate the extent to which this explained variation matters in terms of quantitative impact on outcomes. In addi-

Figure 1
TWO STEPS OF VALUE CREATION
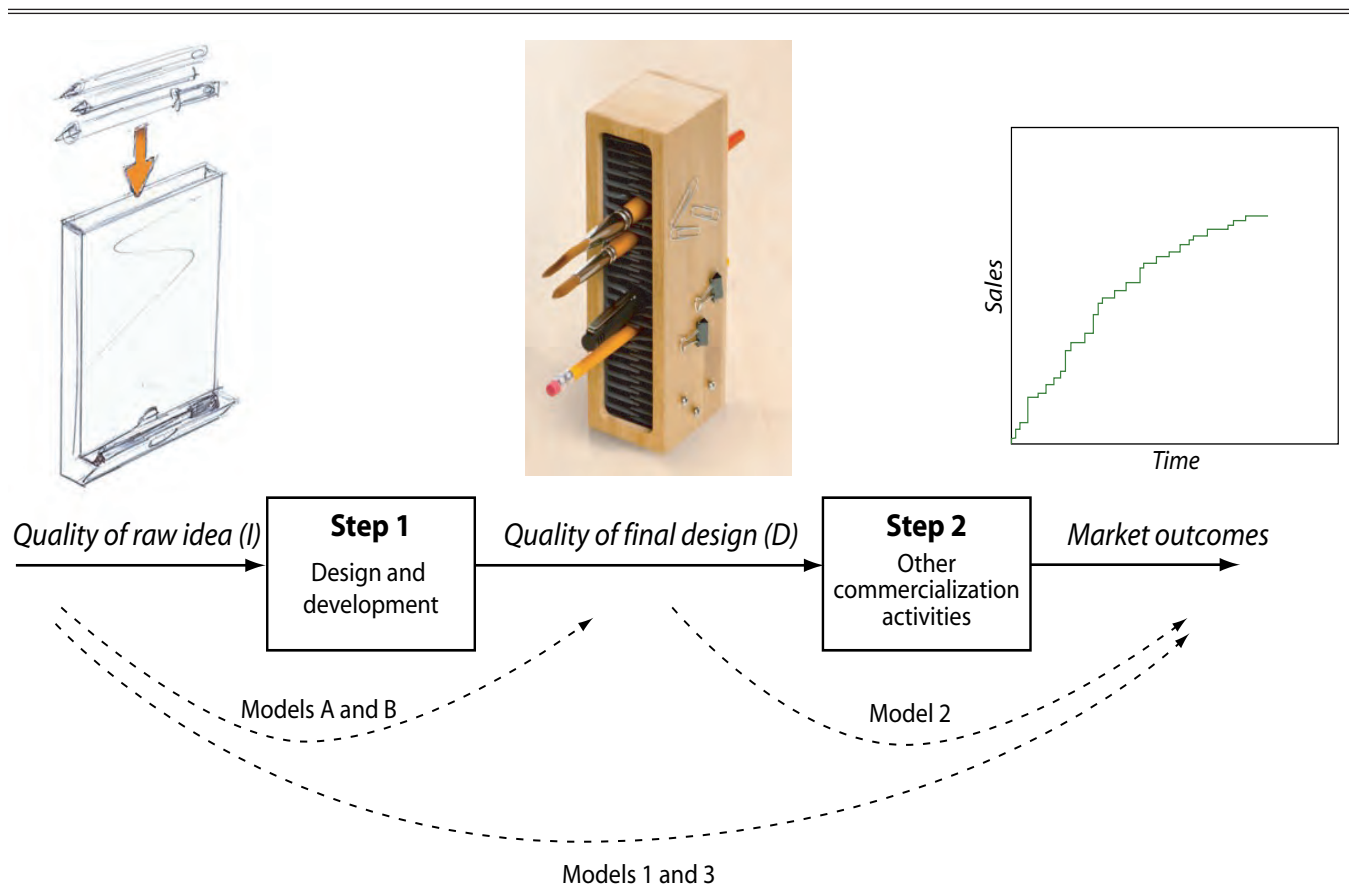
Notes: We analyze the overall relationship between raw idea and market outcomes (Models 1 and 3 in our empirical analysis). We also decompose that relationship into two parts: from raw idea to final design (Models A and B) and from final design to market outcomes (Model 2).

tion, we can also examine the best way to identify good ideas, given that idea quality is an elusive notion.

Empirically, the research question of how much the idea itself matters is challenging for several reasons. First, we face a measurement challenge. Idea quality is a theoretical notion, but we need actual measurements for our analysis. Second, we face a selection challenge. The majority of ideas never receive investment. Thus, we can only observe outcomes for a small fraction of ideas. This data limitation cannot readily be overcome experimentally because of the prohibitive costs associated with commercializing ideas and realizing their outcomes. Third, we face a data availability challenge. We need access to the raw ideas and to the outcomes (e.g., sales results) for a large sample of innovations. However, we need to be able to evaluate the quality of the ideas retrospectively, without polluting that evaluation with knowledge of the actual outcomes.

To address these empirical challenges, we develop a novel data set. We use data from the community-driven product development company, Quirky.com ("Quirky" hereinafter). Quirky runs weekly tournaments on its website, selects the best raw ideas, and leads a product development effort supported by the Quirky community. The successfully developed ideas are sold in an online store on the site. Due to the community orientation of the site, the development process is transparent. The raw ideas are available on the site, and the sales figures for each product are updated as orders come in. We use multiple measures of idea quality: purchase intent measures from a consumer panel and expert ratings of the ideas.

Regarding our central question, our results indicate that ideas are indeed important to sales. We find that the quality of raw ideas, as estimated by commercially feasible techniques, is a statistically significant predictor of market outcomes. The raw idea itself explains only a modest amount of variation in outcomes, but even the modest variation corresponds to an economically important impact on value. We also conclude that for the domain we study, surveys of consumers are a better way to determine a "good" idea than are ratings by even highly experienced experts. We consistently find predictive power in a sample as small as 20 consumers.

We organize the remainder of the article as follows: We first discuss prior research in related areas. We then lay out the conceptual framework; discuss the data, the analysis, and the results; and conclude with a discussion of the managerial implications and our study's limitations.

PRIOR WORK

The current study is one of the few to examine the relationship between the quality of raw ideas and market outcomes. Extant studies of idea generation implicitly assume that better ideas have a significantly positive impact on better market outcomes, suggesting that firms should invest substantial resources in generating better ideas. We do not posit that better ideas may lead to worse market outcomes; rather, we ask how important the quality of the raw idea actually is as a determinant of success. To date, no published studies have empirically examined this question using both the raw ideas as originally proposed and market outcomes. Our study fills that gap. The following subsections detail how our work adds to the existing literature.

Our Unit of Analysis Is Raw Idea as Originally Proposed

Whereas other authors have studied the extent to which attributes and early product evaluations predict market success (e.g., Åstebro 2003; Chandy et al. 2006; Eliashberg, Hui, and Zhang 2007; Goldenberg, Lehmann, and Mazursky 2001; Kamakura, Basuroy, and Boatwritght 2006; Morwitz, Steckel, and Gupta 2007; Rogers 1995), none that we know of have used the raw ideas as originally proposed, and many are based on retrospective descriptions of the products. For example, Goldenberg, Lehmann, and Mazursky's (2001) studies use attributes coded from patents and retrospective synopses of ideas described in books. Chandy et al. (2006) focus on pharmaceuticals and also rely on patents as the basis for describing products. Eliashberg, Hui, and Zhang (2007) address the question of whether movies with certain plot elements (i.e., a model of movie "quality") predict box office results. In their study, the "idea" is a 4- to 20-page narrative summary of the movie, created retrospectively. Similarly, in Study 2 of Morwitz, Steckel, and Gupta (2007), concept tests seem to have predictive validity; however, these "concepts" are descriptions of products ready to be launched. Other forms of market research, such as simulated test markets (e.g., Clancy, Krieg, and Wolf 2006), typically work with elaborate product descriptions. Conceptually, our definition of raw idea quality is similar to this existing literature: raw idea quality is an estimate of how successful an idea would be if developed and sold in the market. The key difference between our research and existing research is that we work with the raw ideas as originally proposed.

Although not based on actual ideas proposed in practice, Dahan et al.'s (2011) securities trading of concepts (STOC) article is one work that relates raw ideas to market outcomes. The researchers generate and describe the ideas in simple visual and verbal depictions based on levels of attributes. In their study of crossover vehicles, the authors examine the relationship between STOC evaluations and market shares and do not find a statistically significant relationship between ideas and outcomes. This provocative finding motivates additional exploration of the research question.

Why is it important to understand the role of the quality of the raw ideas rather than that of more fully developed ideas? Working with raw ideas is a real task for firms, which must sort through dozens, hundreds, or even thousands of possibilities typical of the "fuzzy front end" (Hauser, Tellis, and Griffin 2006) of innovation. Establishing the strength of the relationship between raw ideas and outcomes can help organizations make informed decisions about investments in idea generation and selection.

Our Dependent Variable Is Market Outcomes

Whereas other studies have addressed various questions related to the quality of raw ideas, none that we know of have used market outcomes as a dependent variable. Conceptually, we define market outcomes as how well a product that has been developed and sold performs in the market. Many studies related to idea generation have tracked a dependent variable related to quality of the ideas, often with an evaluation by an expert panel. Goldenberg, Mazursky, and Solomon (1999) use a panel of three senior marketers to

evaluate ideas; Diehl and Stroebe (1987) use one or two research assistants to rate ideas on originality and feasibility; and Dahl and Moreau (2002) use a panel of three experts to judge originality in one study and use 19 consumers in another; they also use panels of 4 or 16 consumers to indicate willingness to pay for raw ideas. Girotra, Terwiesch, and Ulrich (2010) analyze the practices in the academic literature for judging quality of ideas and conclude that the best ways to estimate idea quality are with holistic ratings of business value by trained experts and with purchase intent surveys of consumer panels. Although all these studies use the ideas as originally generated, none of them have an available market metric.

Why is it important to use market outcomes rather than other measures of quality? For both practitioners and scholars, the survey and judgment measures are a cheaper and more convenient proxy for market outcomes. However, the underlying assumption in using nonmarket measures is that they correspond to ultimate business value as would be determined in a market. We test this assumption. We acknowledge that we study only a particular empirical context, but our results provide some guidance about the relationship between widely used premarket quality measures (namely, expert evaluations and consumer surveys) and market outcomes.

Relative Influence of Raw Idea and Final Design

Whereas previous studies have examined the relationship either between refined ideas and market outcomes or between raw ideas and quality judgments, none that we know of have investigated how the stages of actual idea development contribute to value creation. Similar to our definition of raw idea quality, final design quality is an estimate of how successful a particular design for a product will be when it is sold in the market. Our novel data set enables us not only to examine the broader relationship between raw idea and market outcome but also to study two steps: (1) from raw idea to final design and (2) from final design to market outcome. In so doing, we can conclude how much of the uncertainty about market outcomes is resolved with the design and development of the product based on the raw idea.

FRAMEWORK FOR INNOVATION SUCCESS

By its very nature, innovation is a highly uncertain activity. Much has to happen to and around an idea before value is ultimately created. As ideas progress through the new product development process, from idea generation to product launch, various activities and decisions contribute to value creation. In this study, we examine both the role of the quality of the raw idea itself and the role of the development decisions that shape the ultimate product. In this section, we consider how to think about the quality of the raw idea and the role of the design process and how both can be analyzed together.

Defining the Quality of the Raw Idea

The true quality of a raw idea is a theoretical notion. Quality cannot be observed directly. We define the quality of the raw idea as a continuous variable reflecting the expected value of pursuing that idea given the innovator's particular context. Context is important: the quality of the

idea of an "undo button in an elevator" will be higher for Otis Elevator than for IBM. In practice, although the true quality of an idea cannot be observed, it can be estimated. We use ratings of the raw ideas to measure their conceptual and practical quality. For example, a panel of experts can rate the idea, or a survey of consumers can be used to measure purchase intent.

Girotra, Terwiesch, and Ulrich (2010) examine multiple dimensions of idea quality (technical feasibility, novelty, specificity, demand, and overall value). They conclude that multiple dimensions of quality load on a single factor and are highly correlated with holistic assessments and purchase intent measures, establishing that a single readily estimated metric can capture multiple dimensions.

Rust, Moorman, and Dickson (2002) and Golder, Mitra, and Moorman (2012) discuss the concept of quality more generally, beyond idea quality. Our notion of quality relates to the revenue-enhancing activity in Rust, Moorman, and Dickson because it springs from product innovation. For the same reason, it spans the quality production and perception processes described in Golder, Mitra, and Moorman.

Resolution of Uncertainty in the Innovation Process

In most innovation processes, the innovator begins with an idea as well as an estimate of value based largely on his or her knowledge about the perceived quality of that idea. Then, the development process resolves further uncertainty. Finally, the commercialization process is completed and the exogenous factors are realized to reveal the idea's value. Figure 1 depicts this typical sequence. In our analysis, we examine the overall relationship between the quality of the raw idea and market outcomes (the results of both Steps 1 and 2), and we study the two steps separately.

Overall relationship between raw idea and market outcomes. After an idea has been developed into a product and sold, market outcomes can be tracked. Our central questions, then, are as follows: How much of the variation in market outcomes is explained by variation in raw idea quality? How large is the impact of better raw idea quality on market outcomes? Finally, what is a good way to measure raw idea quality?

Role of the development process. In our framework, we also consider an intermediate step, the result of the design and development process that creates the final design sold to consumers. We expect that some of the variation in market outcomes would be explained not only by the raw idea quality but also by what the innovator actually does with the idea. Armed with an estimate of the final design quality, we can decompose the analysis and examine how much of the variance in final design quality is explained by variance in raw idea quality and then determine how much of the variance in market outcomes is explained by variance in final design quality.

DATA

Our data set comprises raw ideas and final designs from a product development and commercialization company, independent evaluations of those ideas and designs, and market outcome measures for the products. We discuss these components in the following subsections.

Company

Quirky is a community product development website. The company specializes in "consumer products that could retail for under $150 and don't involve integrated software," including products for both the home (e.g., kitchen accessories) and the office (e.g., products to keep electronic devices organized). Quirky runs weekly contests in which community members typically contribute more than 100 ideas. The ideas are described with text and/or images, and the best idea(s) are selected from each contest. Moving through the product development process, community members contribute to market research, product design, and naming. Members earn points for participating in the process and then earn money based on those points and product sales. The development is not completely crowdsourced; Quirky employees are heavily involved in the final selection of ideas and the actual development and production arrangements for the products.

The site was attractive as a data source for our research questions because it publicly displayed all the ideas from the contests as well as the sales figures for the products in the store. This transparency springs from the community involvement in the site: community members have an interest in tracking the sales progress of products to which they have contributed. Another attractive feature of the Quirky data is that the single development and sales platform controls the variation in exogenous factors to some extent. There are still exogenous factors that affect outcomes for the products in our sample, but some of those, such as the size of the aware population, are relatively constant for all the products on the site. Alexa.com (2013) reports Quirky's demographics: "Based on internet averages, quirky.com is visited more frequently by users who are in the age range 25–34, have no children, are college educated and browse this site from school."

Ideas

Our data set comprises 160 products from the Quirky store. These products were all offered for sale during one of our two data collection periods, March 2011–November 2011 and December 2012–March 2013. We retrieved ("scraped") from Quirky's site the description of the product in the store and images of the final designs. We collected sales figures on a regular basis during each data collection period.

We also retrieved the "raw idea" (text and images, if available) associated with each product in the store. There were 149 such raw ideas; some of the 160 products were developed from the same raw idea. Figure 1 shows a visual depiction of a raw idea from our data set along with the corresponding final design that was developed from the raw idea; Figure WA1 in the Web Appendix provides the accompanying text descriptions. Two research assistants independently classified the products into categories on the basis of the room of the house for which the product was intended (bathroom, bedroom, garage/utility room, children's room, kitchen, or office). They agreed on 78% of the categories in their independent classifications; they then reconciled the discrepancies. The consensus of our research assistants showed that the 160 products comprised 14 bathroom products, 14 bedroom products, 38 garage/utility room products, 5 children's products, 49 kitchen products, and 40 office products.

Finally, we also generated a random sample of 100 raw ideas from all the idea contests—that is, the entire population of ideas on the site, not only ideas that were developed. (Exactly 1 of these 100 ideas was selected for the store, consistent with the approximately 1% selection ratio in the Quirky contests overall.)

Measuring Market Outcomes

We tracked units sold in the store and recorded prices for each of the products, which enabled us to observe both their sales volume and revenue. Products were introduced to the store continually and, therefore, at different times. Quirky reports not only units sold but also days in the store. The introduction timing raises the question whether we can compare unit sales (or revenue) for a product that has been in the store 30 days with one that has been in the store 330 days. Figure 2 shows a sample of sales trajectories in the data. We use three approaches to address the varied launch dates. First, to normalize for the length of time in the store, we observe sales rates, that is, units sold divided by days in the store. Second, we use the sales trajectories to estimate projected units for each product. We made these projections with S-curve forecasts using the Bass (1969) model (Srinivasan and Mason 1986). Using projected units relaxes some of the assumptions implicit in using sales rates, acknowledging that products are not offered for sale forever and that sales rates may vary throughout the life of a product. To solve the nonlinear optimization problem for each product, we first performed a grid search over the parameter space (the p, q, and m parameters of the Bass model) and then used SAS PROC NLIN to find the best-fitting curve using our grid solution as a starting point. Third, for the models with sales rate as the dependent measure, we ran variations that included a control for the number of days in the store.

Because we do not know actual manufacturing costs, we do not know the exact profit margins of each product. However, the gross margins for a direct-to-consumer specialty retailer of its own proprietary household products are typically high (i.e., greater than 75%), so revenues are very highly correlated with profits (Ulrich and Eppinger 2011).

Furthermore, to address potential endogeneity of price in the estimation, we use estimates of manufacturing costs. The images and descriptions reveal information about the materials, size, number of parts, and types of parts (e.g., whether the product contains electronics) for the final designs, all major drivers of cost. We estimate manufacturing cost on the basis of these factors, as prescribed by Ulrich and Eppinger (2011). We model cost as the sum of materials costs, part processing costs, assembly costs, and transportation costs. Then, we model each of those elements in turn as a function of the product parameters. For example, we determine transportation costs from the product package dimensions and prevailing freight costs between the manufacturing site (China) and the United States. The correlation between the natural log of estimated cost and the natural log of price is .78.

In summary, we track the following outcomes: sales rate, units sold, projected units (from the Bass model), and revenue. These measures are consistent with the results of the Product Development Management Association Taskforce

Figure 2
SALES TRAJECTORIES
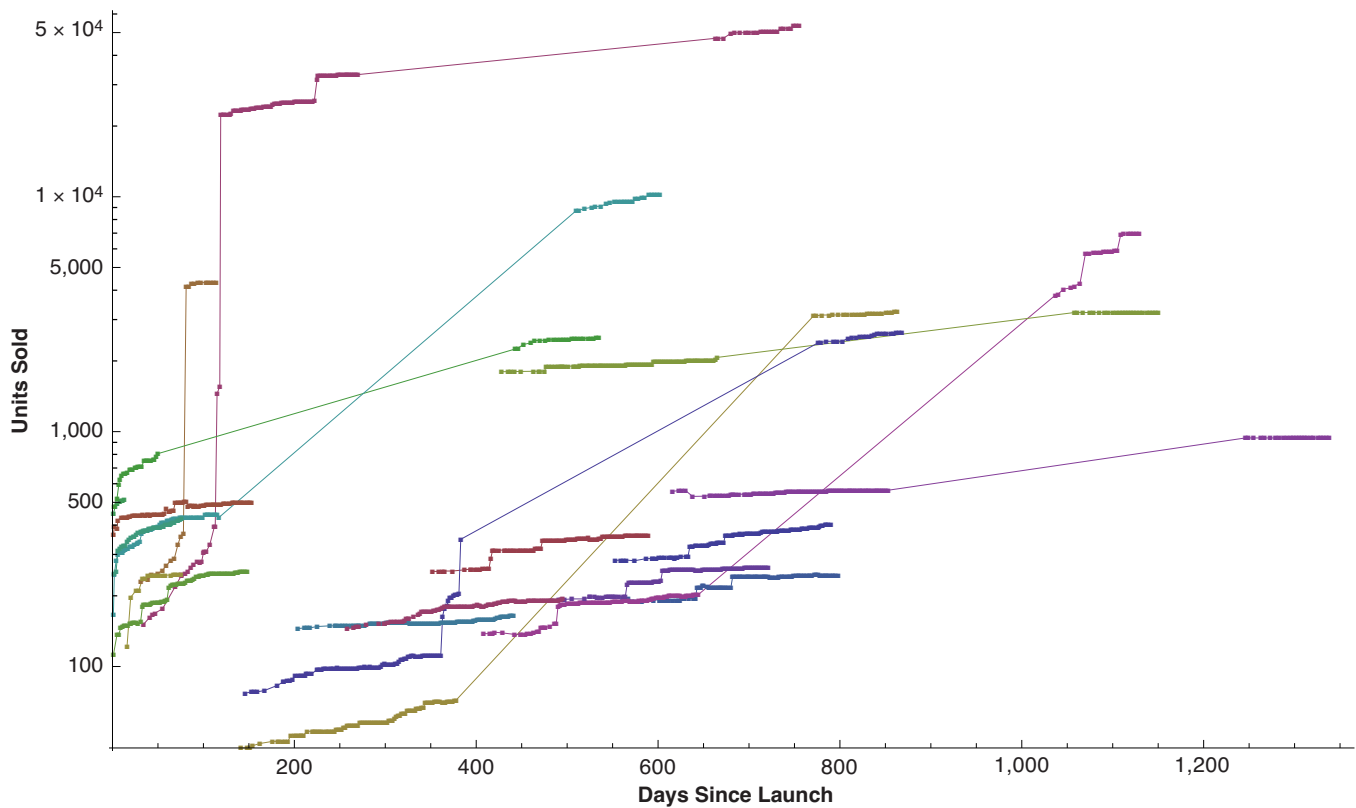
Notes: The sales trajectories shown are for 20 of the products in the sample. Drops in units sold reflect returns. The trajectories shown are the sales during the observation periods from March 25, 2011, to November 17, 2011, and December 14, 2012, to March 16, 2013. The diagonal straight lines interpolate in the noncollection interval; vertical steps represent large orders, likely from retail partners.

on Measures of Success and Failure (Griffin and Page 1993, 1996).

Measuring Quality of the Raw Idea

We measure the quality of the raw idea in two ways. The first measure is purchase intent of the raw idea. Probability of purchase conditional on awareness and availability, for a given price, is essentially a measure of the idea's quality. The better the solution relative to the alternatives and the more pervasive the need it addresses, the more likely a user in the target market is to purchase the innovation. Purchase intent is one of the recommended measures from Girotra, Terwiesch, and Ulrich's (2010) study of idea quality. Moore (1982) documents that concept screening using a purchase intent question is an established industry practice.

We measured purchase intent of the 149 raw ideas that made it into the store and also for the 100 random raw ideas drawn from the entire population of ideas on the site. Paid online panelists from Qualtrics viewed product descriptions (text and, if provided by the originator of the idea, an image) and rated their purchase intentions on a five-point scale (1 = "definitely not," and 5 = "definitely").

We used an attention filter question to screen out people who were not reading the survey. The attention filter was formatted exactly like the concept descriptions in the survey, but in the place of the concept title was "Survey Read-ing Verification," and in place of the concept description was an instruction to select the leftmost option ("definitely not").

We divided the 249 ideas randomly into blocks of 49 or 50, with 29 or 30 ideas from the 149 ideas that made it into the store in each block, and 20 of the 100 random ideas in each block. Each panelist was assigned to a block and shown the ideas from that block in random order. There were 1,438 responses, with each of the five blocks rated by between 282 and 293 panelists. (The block design helped us compare the interrater reliability of these panelists with that of our experts. The unequal numbers resulted from random assignment, screening out, and incompletion.) The Web Appendix presents details on self-reported information from the panelists on gender, age group, and employment status.

We translated the purchase intent responses into a single overall purchase intent score by weighting each response with 0, .25, .5, .75, and 1 (Jamieson and Bass 1989). We present the results using this weighting, and the results hold when the weighted measures are logged. The results are largely insensitive to the weights; for example, they are robust to convex weights on purchase intent responses (.01, .05, .1, .25, .5) found by Haley and Case (1979). Consistent with typical practice for testing raw ideas (for which design concepts have not yet been developed), the product descriptions did not suggest prices. Ottum (2005, p. 295) explains

this common practice: "It is usually a good idea not to put a price on ... early ideas, because the goal of concept testing is to get a read on customer interest in the general idea."

Our main analysis focuses on the ideas that made it into the store, but we also used the purchase intent results for the 100 random ideas. First, we use the standard deviation of purchase intent of the random ideas to estimate the impact of having a one-standard-deviation-better idea. Second, we note that the standard deviation of purchase intent of the 149 ideas that made it into the store is very close to the standard deviation of purchase intent of the 100 random ideas (.0796 vs. .0841). This finding suggests that the results about the relationship between idea quality and outcome are not dramatically skewed by a restricted variance on idea quality in the set of selected ideas compared with the whole population.

The second measure of idea quality was ratings by experts. We used seven experts in consumer products marketing and product development. They rated 98 of the store ideas and 50 of the set of 100 random ideas. (All seven experts reviewed the same 148 ideas.) The 98 ideas came from the initial phase of our data collection, and the 50 ideas were selected randomly from the set of 100 random ideas. Each expert has at least 15 years of experience in designing, developing, or commercializing consumer products. The experts all have experience in multiple product categories throughout the course of their career, with a great deal of overlap in the set of categories in our data set. The qualifications of our experts compare favorably with those reported in the literature (e.g., Dahl and Moreau 2002; Diehl and Stroebe 1987; Girotra, Terwiesch, and Ulrich 2010; Goldenberg, Mazursky, and Solomon 1999).

Experts were asked to rate the ideas on a ten-point scale on the basis of anticipated units sold. The question to the experts was phrased in terms of units sold to make it comparable to the consumer survey. (The purchase intent measure yields an estimate in terms of units.) As with the raw ideas shown to consumers, prices associated with raw ideas would not yet be determined, and we therefore directed the experts to assume that the "resulting products would be priced appropriately."

Measuring Quality of the Final Design

We measured the quality of the final design with purchase intent, using a procedure similar to that for the raw ideas. A different group of paid online panelists viewed product descriptions and images of the final designs, as portrayed in the online Quirky store. They answered a priced purchase intent question using the same five-point scale as described previously. Each panelist rated 53–54 final designs and answered an attention filter question. There were 363 responses, with each of the three blocks rated by between 113 and 129 panelists. The Web Appendix presents details on self-reported information from the panelists on gender, age group, and employment status.

The descriptions of the final designs *did* include prices because Quirky sets product prices as part of the final development process. For most of the products, the actual market prices were constant throughout the observation periods. For the products with price changes, we calculated a representative price, which we defined as the "modal quantity" price, or the price at which the largest quantity was sold. Of

course, we would expect a relationship between price and quantity, but we believed that this approach was the most genuine way of capturing a representative price. This representative price was the price we used in our survey.

One of the challenges in finding a suitable data set for this study is that raters of idea quality need to be unaware of the actual products and their commercial success. Our analysis relies on our finding that the Quirky market is limited and relatively unknown. We could not take the same measures for products sold in mass, mainstream channels such as Target and Wal-Mart; that availability would pollute our ability to go back and measure idea quality. In our purchase intent surveys, we verified that our respondents were not biased by knowledge of market outcomes by asking their familiarity with a set of online retailers that included Quirky.com. Very few (less than 4%) had visited Quirky's website, and we screened out those respondents.

In these measures, we estimated the quality of the raw idea, the quality of the final design, and the market outcome, the essential ingredients for addressing our research questions. Table 1 summarizes the descriptive statistics of our measures, and Table 2 shows the correlation table.

ANALYSIS AND RESULTS

In this section, we present the answers to our research questions. Does the quality of the raw idea matter? If so, how much? And what is the best way to measure that quality?

We address our central question about the importance of the raw idea by estimating the relationship between measures of raw idea quality and market outcomes. Our data enable us not only to test that start-to-finish relationship but also to decompose the analysis into two steps: (1) the relationship between raw idea quality and purchase intent of the final design and (2) the relationship between purchase intent of the final design and market outcomes (see Figure 1).

To analyze the relationship between the raw idea quality I and the market outcome, the arrow spanning both steps in Figure 1, we estimate Model 1. In this equation, PR represents a control variable for price and CAT represents a vector of dummy variables for product category. For concrete-

Table 1
DESCRIPTIVE STATISTICS

	M	Mdn	SD	N
Price (in US$)	30.12	24.99	24.83	160
Units sold	17,257	490	66,124	160
Projected units	192,834	705	1,964,559	160
Sales rate (units/day)	37.80	3.27	91.03	160
Revenue (in US$)	334,749	12,815	1,519,439	160
Days in store	462.1	400.0	360.3	160
Cost (in US$)	6.29	4.35	6.44	160
Purchase intent of raw idea (0–1), developed ideas	.45	.46	.08	149
Purchase intent of raw idea (0–1), random ideas	.40	.40	.08	100
Purchase intent of final design (0–1)	.28	.27	.08	160
Expert rating (0–10), developed ideas	4.26	4.29	1.22	98
Expert rating (0–10), random ideas	3.37	3.21	1.21	50

Table 2

CORRELATION MATRIX

	Ln(Price)	Ln(Units Sold)	Ln(Projected Units)	Ln(Sales Rate)	Ln(Revenue)	Ln(Days in Store)	Ln(Cost)	Purchase Intent (Raw Idea)	Purchase Intent (Final Design)	Average Expert Rating
Ln(Price)	—									
Ln(Units Sold)	−.35***	—								
Ln(Projected Units)	−.29***	.89***	—							
Ln(Sales Rate)	−.35***	.85***	.73***	—						
Ln(Revenue)	−.02	.94***	.84***	.78***	—					
Ln(Days in Store)	−.01	.31***	.32***	−.24***	.33***	—				
Ln(Cost)	.78***	−.44***	−.38***	−.40***	−.19**	−.08	—			
Purchase intent (raw idea)	.02	.23***	.17**	.25***	.26***	−.03	.06	—		
Purchase intent (final design)	−.45***	.30***	.19**	.36***	.16**	−.09	−.21***	.55***	—	
Average expert rating	−.09	.18*	.14	.24**	.17*	−.14	.04	.49***	.31***	—

*p < .10.
**p < .05.
***p < .01.
Notes: Correlation coefficients for the variables are for ideas that were developed and sold in the store. Correlations with average expert ratings contain only 109 observations, compared with 160 observations for the others.

ness, the dependent variable shown is Ln(Sales Rate), but we also perform the analysis for other outcome measures such as units sold and projected units.

$$(1) \quad Ln(Sales\ Rate) = \beta_0 + \beta_1 I + \beta_2 Ln(PR) + \beta_3 CAT + \varepsilon.$$

Price is a natural control to use, and we expect a negative relationship between the price of the product and the quantity sold. To address the potential endogeneity of price in this model, we use an instrumental variable: estimated manufacturing costs. Our main results use that instrumental

variable in a two-stage least squares (2SLS) procedure. In 2SLS, we first estimate price as function of cost and the other exogenous variables (Stage 1, included in Web Appendix Table WA1) and then use the estimated price in the regression for sales rate (Stage 2, reported in Table 3). Table 3 also shows the results of the Hausman test for endogeneity for each model. When that test shows significance, as it does for the majority of the models, the 2SLS estimates are preferred to ordinary least squares (OLS). We include the analogous OLS analyses in Appendix A.

Table 3

SALES AND FINAL DESIGN MODELS

	A: Dependent Variable: Ln(Sales Rate)					
	Raw Idea→Sales Model 1		Final Design→Sales Model 2		Experts→Sales Model 3	
Constant	3.501***	(1.231)	4.247***	(1.371)	2.975*	(1.594)
Estimated Ln(Price)	−1.330***	(.246)	−1.155***	(.274)	−1.212***	(.314)
Purchase intent (raw idea)	4.885***	(1.810)				
Purchase intent (final design)			3.591*	(2.115)		
Average expert rating					.364**	(.141)
Controls for category	Yes		Yes		Yes	
R²	.32		.31		.34	
N	160		160		109	
Hausman Test (Test of Endogeneity)						
First-stage residuals	.645	(.415)	.954**	(.474)	.725	(.549)

	B: Dependent Variable: Purchase Intent of Final Design			
	Raw Idea→Final Design Model A		Experts→Final Design Model B	
Constant	.138***	(.040)	.282***	(.060)
Estimated Ln(Price)	−.036***	(.008)	−.029**	(.012)
Purchase intent (raw idea)	.551***	(.059)		
Average expert rating			.020***	(.005)
Controls for category	Yes		Yes	
R²	.52		.33	
N	160		109	
Hausman Test (Test of Endogeneity)				
First-stage residuals	−.049***	(.013)	−.050**	(.020)

*p < .10.
**p < .05.
***p < .01.
Notes: Results are from the second stage of 2SLS regressions. Standard errors are in parentheses.

Product categories are another potential source of variation in outcomes (e.g., the degree of need may differ across product categories), so we control for that as well. This control tells us, for example, whether similarly priced kitchen products tend to sell more (or less) than office products. The results are not substantially different when we control only for price and not category.

Throughout our analyses, we use the natural log of the sales rate and of the price: these two variables have values that extend over more than a factor of ten, with a long right tail. We report results with the nonlogged versions of the idea quality measures. The results are not substantially different from the natural logs of those quantities, either.

Purchase Intent

In this subsection, we present the results using purchase intent as the measure of idea quality. We observe from the simple correlation in Table 2 that purchase intent of the raw idea is a statistically significant predictor of the natural log of sales rate ($r = .25$): higher-rated raw ideas predict higher-performing products.

Yet how much does the idea matter? We can answer that question in two ways: variance explained and size of the effect on outcomes. The squared simple correlation and OLS partial R-squared measures[1] show that the raw idea accounts for approximately 5% or 6% of the variance in the natural log of sales rate. Therefore, in terms of variance explained, the effect of the raw idea seems modest but is still statistically significant. In terms of impact, though, this small explained variance translates to a large effect on outcomes. The standard deviation of purchase intent of the raw idea in the random sample of ideas is .0841. In Model 1, the coefficient on purchase intent of the raw idea is 4.885. A one-standard-deviation change in purchase intent translates to a change in the natural log of sales rate estimated at .411 ($= .0841 \times 4.885$). In other words, a one-sigma-better idea corresponds to a 51% increase in sales rate. (The 51% = $e^{.411} - 1$.) Although much of the variance is unexplained, even 5% of the variance in the natural log of sales rate corresponds to a large economic effect.

That analysis applies to the conjunction of the two steps in Figure 1. Next, we turn our attention to each step individually. In Model A, we examine Step 1 from Figure 1. We test whether the purchase intent of the raw idea predicts the purchase intent of the final design by estimating models analogous to the tests of Model 1, but with the dependent variable as the purchase intent of the final design.

(A)   PI-Final Design = $\beta_0 + \beta_1 I + \beta_2 Ln(PR) + \beta_3 CAT + \epsilon$.

Again, we find significance of the raw idea. The squared simple correlation and OLS partial R-squared values of this model exceed .30. The purchase intent of the raw idea is a strong indicator of the purchase intent of the final design.

In Model 2, we examine Step 2 from Figure 1. We test whether the purchase intent of the final design predicts sales rate by estimating models analogous to Model 1 but with an independent variable of the purchase intent of the final design, D.

(2)   Ln(Sales Rate) = $\beta_0 + \beta_1 D + \beta_2 Ln(PR) + \beta_3 CAT + \epsilon$.

In Model 2, the significance of the final design quality operates through price; a negative and significant coefficient in the price estimation stage (Table WA1 in the Web Appendix) shows that higher intent ratings are associated with lower prices. The survey for purchase intent of the final design included the price in the description of the product, and the correlation between purchase intent of the final design and the natural log of price is –.45.

In comparing Figure 1's Step 1 with Step 2 by examining Models A and 2, we observe that the R-squared of the first step (.52) is greater than that of the second step (.31). The OLS partial R-squared values also show the contrast between Steps 1 and 2: controlling for price and category, purchase intent of the raw idea explains 38% of variance in purchase intent of the final design, but purchase intent of the final design only explains 3.7% of the variance in the natural log of sales rate.

These results are robust to other outcome measures and other approaches to control for the different lengths of time products spent in the store. Tables WA2–WA4 in the Web Appendix show, respectively, the results when the outcome variable is units sold, Bass model projections on the whole data set (N = 160), and Bass model projections for products in data set that have been in the store more than six months (N = 109). Imposing a minimum sales history increases the predictive power of purchase intent for sales rate. For example, the R-squared for Model 1 increases from 32% to 40% when the sample is limited to the 76 products that have been in the store for more than 450 days. The estimated coefficients on purchase intent, and thus the size of the effect on sales, also steadily increase with longer sales history. This pattern suggests that purchase intent improves as a predictor of market outcomes when the outcome measure is a longer-term metric.

These strengthened results on subsamples of the data suggest that controlling for days in the store will add explanatory power to the model. Indeed it does, and we provide the results of such an analysis in Table WA5 in the Web Appendix. The magnitudes of the coefficients on the quality measures remain relatively stable compared with those in Table 3.

Our conclusions about purchase intent and outcomes are as follows. First, the quality of the raw idea as measured by purchase intent is a statistically significant predictor of outcomes. Second, even though the percentage of variance explained is modest, the impact on outcomes is large. Third, there is a stronger predictive link between a better raw idea and better final design than there is between a better final design and better outcomes. Finally, outcomes measured with more sales history are better explained by raw idea quality than those with less sales history.

Expert Ratings

Next, we examine estimates of idea quality based on the ratings from our seven experts. We estimate versions of Models 1 and A with average expert ratings as the measure of idea quality.

---

[1]The R-squared values from the 2SLS procedure are not, strictly speaking, interpretable as percentage of variance explained, although they still have the form of such a measure (1 – sum of squared errors/total sum of squares; Wooldridge 1999). The R-squared values from both 2SLS and OLS are close in all of our models, with the 2SLS R-squared values slightly lower; Appendix A reports the OLS R-squared values and the OLS partial R-squared values for the quality variable.

The experts exhibited a relatively low level of agreement with one another. Appendix B shows the correlation matrix for the seven experts and their average. The correlation for the average of the experts with natural log of sales rate is higher than any individual expert. Poor agreement suggests that there will be poor interjudge reliability. Indeed, common measures of interjudge reliability confirm that suggestion (Cronbach's [1951] alpha = .46, Krippendorff's alpha = .085 [Hayes and Krippendorff 2007]). Despite the low level of agreement and the poor reliability, the average of the experts' ratings is a statistically significant predictor of market outcomes, as shown in Models 3 and B in Table 3.

Examining Table 3, it would seem that the experts' predictive power for sales outcomes is comparable to that of the purchase intent of the idea for the consumer panel. Compare the R-squared values in Model 1 (with purchase intent of the raw idea; .32) with Model 3 (with experts; .34). The increase in sales rate for a one-standard-deviation-better idea is also similar, estimated at 55% ($= e^{(1.21 \times .364)} - 1$) for Model 3. (The corresponding values for the purchase intent of the raw idea, reported in the previous subsection, is 51%.)

What Table 3 does not show is sensitivity to other dependent measures. The predictive power of the consumers' purchase intent ratings is largely robust to using units sold and projected units, but the same is not true of the experts' ratings. With these other dependent measures, the significance of the coefficient on the experts' average is marginal (with units sold, see Table WA2 in the Web Appendix) or not significant (with the projected units, see Tables WA3 and WA4 in the Web Appendix).

The contrast between experts' and consumers' predictive power is also evident when we run the models only on the subset of ideas the experts rated, corresponding to N = 109 products.[2] In those comparisons, Model 1 yields an R-squared of .40 (not shown in Table 3) versus Model 3's R-squared of .34 (as shown in Table 3). One explanation for the increased predictive power on the smaller sample is that these products have a longer sales history, on average, than the full sample. As we discussed in the previous subsection, restricting the analysis to products with longer sales histories in our data typically increases the predictive power of the models.

Another discrepancy in comparing experts and consumers is their number. We would like to know how large a set of consumers would need to be to generate equivalent predictive power to that of our seven experts. To identify the expert-equivalent sample size, we pulled repeated subsamples from the consumer respondents, calculated the weighted purchase intent of the sample, and ran the regression models. In the results reported next, we iterated 500 times (i.e., we pulled 500 subsamples for each sample size

we tried). A sample of ten consumer respondents means that we pulled ten respondents from each of the five blocks so that each idea was rated exactly ten times.

Figure 3, Panel A, shows the average R-squared across the 500 trials for eight sample sizes (1, 4, 7, 10, 15, 20, 30, and 50) for Model 1 with the set of 109 products whose ideas were rated by experts. With a sample size of 4, the average R-squared for Model 1 is .33, directly comparable to the experts' .34 (Model 3 in Table 3).

Figure 3, Panel B, shows the percentage of the 500 iterations that yielded a significant coefficient on the purchase intent of the raw idea in Model 1. In 52% of the 500 trials, the purchase intent calculated from only 4 consumers significantly predicts the outcome measure (natural log of sales rate) at $p < .05$ (and in 28% of the 500 at $p < .01$). For as few as 15 consumers, we observe more than a 90% chance of having their collective voice be a significant predictor ($p < .05$) of outcomes. With any 20 consumers from our sample, we were virtually guaranteed to observe significance at $p < .05$.

We draw two conclusions from these data. First, asking consumers to state purchase intentions of raw ideas predicts market outcomes better than does asking experts to predict sales. Consumers are an even more attractive option for firms when considering the higher cost of enlisting experts. Sec-

Figure 3
EVEN SMALL CONSUMER SUBSAMPLES ARE PREDICTIVE

A: Average R-Squared over 500 Iterations

B: Percentage of the 500 Iterations with Significant Predictive Effect
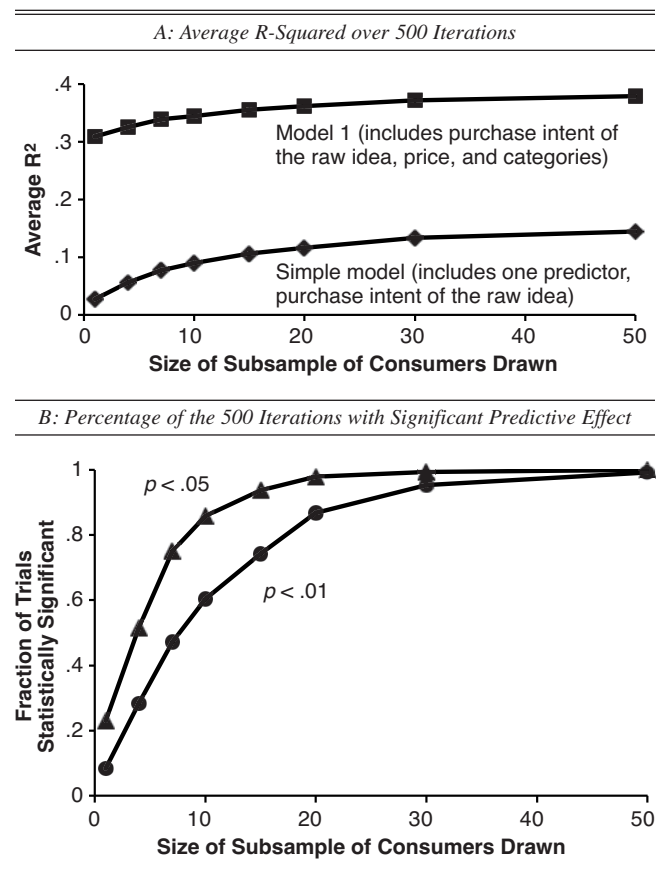
Notes: Panel A shows the average R-squared over 500 iterations as a function of the size of the consumer subsamples drawn. Panel B shows the percentage of the 500 iterations that showed a significant predictive effect of the purchase intent of the raw idea in Model 1 as a function of the size of the consumer subsamples drawn.

[2]To elaborate on the discrepancy in the samples for the experts and the consumers, the difference resulted from the timing and progress of the study. We conducted our first round of data collection in 2011, and our experts only rated the ideas in the first round. When we did a second round of data collection to update the sales figures and collect consumer ratings of all the ideas (including ideas added since our first collection), we did not re-collect from the experts for two reasons. First, unlike the consumer panel, which is managed by a vendor (Qualtrics) from which we can purchase respondents, the experts are specific people who had already expended considerable time and effort for our study. Second, the data from the first round indicated that experts' ratings were not predictive of market outcomes, so re-collecting data from them did not seem a worthwhile use of their time and goodwill.

ond, the consumer results are extremely robust in terms of likelihood of significant prediction and explanatory power, even at small sample sizes. The answer to the question of how many consumers are needed to be equivalent predictors to our seven experts is approximately 4–20: with only 4 consumers, on average, the predictive power is the same as the expert group, and with 20, the predictive power is greater than that of experts, with near certainty of significance of the ratings. Firms do not typically use this type of consumer research at such an early stage, perhaps partly for reasons of secrecy. Our evidence shows that market wisdom is embedded even in small crowds (Surowiecki 2004). Ideas can reliably be vetted early, cheaply, and with limited dissemination.

Gross's (1972) model shows that the reliability, or signal-to-noise ratio, of the screening instrument is an important determinant of value in a creative process. The potency of small samples of consumers suggests that the noise in a consumer purchase intent survey is relatively low, making these types of surveys a useful tool in creating value.

DISCUSSION

In addition to addressing our central research question—essentially a scientific question about how much the idea matters to market success—our results have practical implications for product development processes, particularly the "fuzzy front end" (Hauser, Tellis, and Griffin 2006). We discuss the implications and limitations of our study in the following subsections.

Managerial Implications

The first implication is that ideas matter. Conventional wisdom among some entrepreneurs is that the idea does not matter: ideas are rarely novel, but truly great execution is rare, so the team and its execution ability are the origin of product success. At the root of this belief are two notions. First is the understanding that it is difficult to judge the quality of an idea. Second is that the conjunctive nature of success limits the influence of the idea; in other words, the idea *and* the design *and* the marketing *and* the market conditions all matter. We agree that measurements of idea quality are noisy and that idea quality is only part of the picture, but our results show that idea quality does indeed predict outcomes. Even though it is difficult to know which ideas will be successful, the idea itself still matters.

The second implication pertains to the value of idea selection. Because a good idea leads to a greater level of success, there is value in accurate selection. Dahan and Mendelson (2001), who derive the optimal number of concepts to test, assume early stage evaluations to be unbiased, if noisy, estimates of value. Even if the early stage evaluations are unbiased estimates of true quality, there is a loss in value associated with the noise that results from investing in ideas that appear to be the best ones but actually are not. These results suggest that higher-fidelity screening in the earlier stages may be worth the investment. Our surveys yield predictive information about market outcomes, but other approaches warrant further research. Approaches that simulate markets, such as Dahan et al.'s (2011) STOC, or test markets with actual purchases (if the prototyping economics allow) will help firms determine whether they possess silk or sows' ears in the first place.

The third implication pertains to how to measure idea quality. In our analysis, ordinary consumers' stated purchase intentions were a better gauge of market outcomes than a panel of experts. This finding builds on previous research such as Hoch (1988), who shows that experts have insufficient knowledge about the activities, interests, and opinions of American consumers; Tetlock (2005), who conducts a long-term study of the predictive power of experts; Faulkner and Corkindale (2009); and Ozer (2009). Our experts are product designers, marketers, and consultants with years (and in some cases, decades) of experience. The low level of agreement between experts (Kamakura, Basuroy, and Boatwright 2006) hints that individual experts or even a panel of experts are insufficiently prescient. We find that there is wisdom in a crowd of consumers. Collecting this type of information is currently so inexpensive that not doing so seems foolish.

Caveats and Limitations

Several caveats and limitations bound the implications of our results. The domain of consumer housewares is a multibillion-dollar industry and thus is economically significant in its own right. However, we would not expect the specific numerical results from this domain to apply in, say, film or pharmaceutical industries. This is a fundamental limitation of most empirical research. For example, Chandy et al.'s (2006) empirical study of the pharmaceutical industry offers broad conceptual insights into conversion ability, even though their specific numerical results cannot be generalized to other industries.

The Quirky data are idiosyncratic for other reasons as well. The open nature of the Quirky development process is what allows us to study it. Some of Quirky's customers are also participants in its development process. Understandably, if 100 people developed a sense of purpose and community around the creation of a new kitchen implement, they might also buy it when available for sale and possibly stimulate purchase by others. Bayus (2013) explores some of the idiosyncrasies of an idea crowdsourcing platform. We conjecture that, if anything, the Quirky system limits the variance in the exogenous factors. If so, we would expect that the variance in outcomes explained by the quality of the raw ideas would be greater for Quirky than for firms with more channels of distribution.

In this study, our data do not allow us to deconstruct all the drivers of value. In particular, we do not know whether better ideas attract more talented developers or are promoted more heavily, and therefore, we cannot say whether those factors contribute to better performance. On the one hand, this is a gap in our understanding of value creation in innovation. On the other hand, we can state that if these forces were at work in the setting we studied, overall they did combine in such a way that ideas rated as better by consumers and experts had better market outcomes.

Note that we treat price differently in the two steps (as shown in Figure 1): in surveying consumers and experts about the raw ideas, we did not include prices, and yet we did include prices in surveying consumers about the final designs. We chose that information structure because it represents knowledge that companies would actually have at each step. In a future study, we could further control for the effect of price in the comparison of the steps by collecting a version of the purchase intent of the raw idea that includes

the prices, because the stated prices do seem to have a strong negative effect on stated purchase intent.

We attempted to match the profile of the purchase intent survey respondents with the profile of Quirky's customers. Inevitably, that matching is inexact. A sample that better fits Quirky's customer population might result in purchase intent measures that are better outcome predictors than the ones we obtained.

As we noted in our "Data" section, we directly observe only revenue, not profit. We have addressed this issue with a cost model, but actual observations of cost would reduce the noise in the costs and ensure that they were unbiased. We also used several dependent measures because we do not have the ideal measure of long-term incremental profit.

We have suggested that Quirky aims to select the best ideas to develop, and we have considered the success of each idea individually. In other words, we have not considered portfolio effects, which would add a dependency among the successes of the projects. We ignored any portfo-

lio effects for three reasons. First, the Quirky products are from such a broadly defined product space that there seems to be little risk of cannibalization. Second, there is no evidence in its stated criteria—the five areas of community, staff, design, market, and viability—that Quirky considers such effects. Third, measuring incremental value to a given portfolio seems to be prohibitive practically. Conceptually, the question of portfolio composition is a worthwhile one, but it does not seem to be a key issue in this setting.

It could be the case that experts are better at judging the relative promise of ideas in a narrow scope than in a broad one. For example, Dahl and Moreau (2002) have participants generate concepts as solutions for the commuting diner. Perhaps the ideas are easier to discriminate because they are easier to compare. However, the opposite could be true: given that the ideas are more similar (Kornish and Ulrich 2011), they could be more difficult to discriminate. Our work suggests that further investigation of the role of experts is warranted.

Appendix A
OLS SALES AND FINAL DESIGN MODELS

A: Dependent Variable: Ln(Sales Rate)					
	Raw Idea→Sales Model 1		Final Design→Sales Model 2		Experts→Sales Model 3
Constant	2.838*	(1.149)	2.928*	(1.191)	2.070 (1.431)
Ln(Price)	−1.108**	(.198)	−.847**	(.224)	−.978** (.258)
Purchase intent (raw idea)	4.839**	(1.803)			
Purchase intent (final design)			4.856*	(2.003)	
Average expert rating					.378** (.140)
Controls for category	Yes		Yes		Yes
N	160		160		109
R²	.33		.32		.34
Adjusted R²	.29		.29		.30
Partial R² (idea quality)	.045		.037		.067

B: Dependent Variable: Purchase Intent of Final Design				
	Raw Idea→Final Design Model A		Experts→Final Design Model B	
Constant	.189**	(.037)	.344**	(.053)
Ln(Price)	−.053**	(.006)	−.045**	(.010)
Purchase intent (raw idea)	.554**	(.058)		
Average expert rating			.019**	(.005)
Controls for category	Yes		Yes	
N	160		109	
R²	.54		.35	
Adjusted R²	.52		.30	
Partial R² (idea quality)	.38		.11	

*p < .05.
**p < .01.
Notes: Results of OLS regressions for Models 1, 2, 3, A, and B are as shown in Figure 1. Standard errors are in parentheses.

Appendix B
CORRELATION MATRIX FOR EXPERTS

	Expert 1	Expert 2	Expert 3	Expert 4	Expert 5	Expert 6	Expert 7	Average
Expert 1	—							
Expert 2	.27	—						
Expert 3	.08	.12	—					
Expert 4	.32	.13	.02	—				
Expert 5	.13	.24	.05	.12	—			
Expert 6	.09	.20	.05	.23	.16	—		
Expert 7	.18	.14	.12	.01	−.03	.03	—	
Average	.53	.56	.33	.51	.60	.49	.43	—
Ln(Sales Rate)	−.02	.19	−.04	.08	.22	.16	.11	.24

Notes: This table shows a correlation matrix of expert ratings on 148 raw ideas as well as correlations with the natural log of sales rate for the 109 ideas that were launched. Experts rated 98 of the raw ideas that were developed into products and 50 of the randomly selected products. All seven experts rated the same 148 ideas.

REFERENCES

Alexa.com (2013), "Quirky.com Site Info," (accessed October 18, 2013), [available at http://www.alexa.com/siteinfo/quirky.com#].

Arthur D. Little Inc. (1921), "Report: On the Making of Silk Purses from Sows' Ears," MIT Institute Archives and Special Collections, (accessed October 18, 2013), [available at http://libraries.mit.edu/archives/exhibits/purse/].

Åstebro, Thomas (2003), "The Return to Independent Invention: Evidence of Unrealistic Optimism, Risk Seeking or Skewness Loving?" *The Economic Journal*, 113 (484), 226–39.

Bass, Frank M. (1969), "A New Product Growth Model for Consumer Durables," *Management Science*, 15 (1), 215–27.

Bayus, Barry L. (2013), "Crowdsourcing New Product Ideas over Time: An Analysis of the Dell IdeaStorm Community," *Management Science*, 59 (1), 226–44.

Chandy, Rajesh, Brigette Hopstaken, Om Narasimhan, and Jaideep Prabhu (2006), "From Invention to Innovation: Conversion Ability in Product Development," *Journal of Marketing Research*, 43 (August), 494–508.

Clancy, Kevin J., Peter C. Krieg, and Marianne M. Wolf (2006), *Market New Products Successfully*. Lanham, MD: Lexington Books.

Cronbach, Lee J. (1951), "Coefficient Alpha and the Internal Structure of Tests," *Psychometrika*, 16 (3), 297–334.

Dahan, Ely, Adlar J. Kim, Andrew W. Lo, Tomaso Poggio, and Nicholas Chan (2011), "Securities Trading of Concepts (STOC)," *Journal of Marketing Research*, 48 (June), 497–517.

—— and Haim Mendelson (2001), "An Extreme Value Model of Concept Testing," *Management Science*, 47 (1), 102–116.

Dahl, Darren W. and Page Moreau (2002), "The Influence and Value of Analogical Thinking During New Product Ideation," *Journal of Marketing Research*, 39 (February), 47–60.

Diehl, Michael and Wolfgang Stroebe (1987), "Productivity Loss in Brainstorming Groups: Toward the Solution of a Riddle," *Journal of Personality and Social Psychology*, 53 (3), 497–509.

Eliashberg, Jehoshua, Sam K. Hui, and Z. John Zhang (2007), "From Story Line to Box Office: A New Approach for Green-Lighting Movie Scripts," *Management Science*, 53 (6), 881–93.

Faulkner, Margaret and David R. Corkindale (2009), "Are Experts Better Than Potential Users in Predicting the Uptake of an Innovation? Extending the Use of the Juster Scale," *Technological Forecasting & Social Change*, 76 (7), 910–16.

Girotra, Karan, Christian Terwiesch, and Karl T. Ulrich (2010), "Idea Generation and the Quality of the Best Idea," *Management Science*, 56 (4), 591–605.

Goldenberg, Jacob, Donald R. Lehmann, and David Mazursky (2001), "The Idea Itself and the Circumstances of Its Emergence as Predictors of New Product Success," *Management Science*, 47 (1), 69–84.

——, David Mazursky, and Sorin Solomon (1999), "Toward Identifying the Inventive Templates of New Products: A Channeled Ideation Approach," *Journal of Marketing Research*, 36 (May), 200–210.

Golder, Peter N., Debanjan Mitra, and Christine Moorman (2012), "What Is Quality? An Integrative Framework for Processes and States," *Journal of Marketing*, 76 (July), 1–23.

Griffin, Abbie and Albert L. Page (1993), "An Interim Report of Measuring Product Development Success and Failure," *Journal of Product Innovation Management*, 10 (4), 291–308.

—— and —— (1996), "PDMA Success Measurement Project: Recommended Measures for Product Development Success and Failure," *Journal of Product Innovation Management*, 13 (6), 478–96.

Gross, Irwin (1972), "The Creative Aspects of Advertising," *Sloan Management Review*, 14 (1), 83–109.

Haley, Russell I. and Peter B. Case (1979), "Testing Thirteen Attitude Scales for Agreement and Brand Discrimination," *Journal of Marketing*, 43 (October), 20–32.

Hauser, John, Gerard J. Tellis, and Abbie Griffin (2006), "Research on Innovation: A Review and Agenda for Marketing Science," *Marketing Science*, 25 (6), 687–717.

Hayes, Andrew F. and Klaus Krippendorff (2007), "Answering the Call for a Standard Reliability Measure for Coding Data," *Communication Methods and Measures*, 1 (1), 77–89.

Hoch, Stephen J. (1988), "Who Do We Know: Predicting the Interests and Opinions of the American Consumer," *Journal of Consumer Research*, 15 (3), 315–24.

Jamieson, Linda F. and Frank M. Bass (1989), "Adjusting Stated Intention Measures to Predict Trial Purchase of New Products: A Comparison of Models and Methods," *Journal of Marketing Research*, 26 (August), 336–45.

Johnson, Steven (2010), *Where Good Ideas Come From: The Natural History of Innovation*. New York: Riverhead Books.

Kamakura, Wagner A., Suman Basuroy, and Peter Boatwright (2006), "Is Silence Golden? An Inquiry into the Meaning of Silence in Professional Product Evaluations," *Quantitative Marketing and Economics*, 4 (2), 119–41.

Kornish, Laura J. and Karl T. Ulrich (2011), "Opportunity Spaces in Innovation: Empirical Analysis of Large Samples of Ideas," *Management Science*, 57 (1), 107–128.

Moore, William L. (1982), "Concept Testing," *Journal of Business Research*, 10 (3), 279–94.

Morwitz, Vicki G., Joel H. Steckel, and Alok Gupta (2007), "When Do Purchase Intentions Predict Sales?" *International Journal of Forecasting*, 23 (3), 347–64.

Ottum, Brian D. (2005), "Quantitative Market Research," in *The PDMA Handbook of New Product Development*, 2d ed., Kenneth B. Kahn, ed. New York: John Wiley & Sons, 279–301.

Ozer, Muammer (2009), "The Roles of Product Lead-Users and Product Experts in New Product Evaluation," *Research Policy*, 38 (8), 1340–49.

Ries, Eric (2011), *The Lean Startup: How Today's Entrepreneurs Use Continuous Innovation to Create Radically Successful Businesses*. New York: Crown Business.

Rogers, Everett M. (1995), *Diffusion of Innovations*. New York: The Free Press.

Rust, Roland T., Christine Moorman, and Peter R. Dickson (2002), "Getting Return on Quality: Revenue Expansion, Cost Reduction, or Both?" *Journal of Marketing*, 66 (October), 7–24.

Srinivasan, V. and Charlotte H. Mason (1986), "Nonlinear Least Squares Estimation of New Product Diffusion Models," *Marketing Science*, 5 (2), 169–78.

Surowiecki, James (2004), *The Wisdom of Crowds*. New York: Random House.

Terwiesch, Christian and Karl T. Ulrich (2009), *Innovation Tournaments: Creating and Selecting Exceptional Opportunities*. Boston: Harvard Business Press.

Tetlock, Philip E. (2005), *Expert Political Opinion, How Good Is It? How Can We Know?* Princeton, NJ: Princeton University Press.

Ulrich, Karl T. and Steven D. Eppinger (2011), *Product Design and Development*, 5th ed. New York: McGraw-Hill Higher Education.

Wooldridge, Jeffrey M. (1999), *Introductory Econometrics: A Modern Approach*. Mason, OH: South-Western.