

THE ORIGIN OF PLUTO'S ORBIT: IMPLICATIONS FOR THE SOLAR SYSTEM BEYOND NEPTUNE

RENU MALHOTRA

Lunar and Planetary Institute, 3600 Bay Area Blvd., Houston, Texas 77058

Electronic mail: renu@lpi.jsc.nasa.gov

Received 1995 February 22; revised 1995 April 3

ABSTRACT

The origin of the highly eccentric, inclined, and resonance-locked orbit of Pluto has long been a puzzle. A possible explanation has been proposed recently [Malhotra, 1993, *Nature*, 365, 819] which suggests that these extraordinary orbital properties may be a natural consequence of the formation and early dynamical evolution of the outer solar system. A resonance capture mechanism is possible during the clearing of the residual planetesimal debris and the formation of the Oort Cloud of comets by planetesimal mass loss from the vicinity of the giant planets. If this mechanism were in operation during the early history of the planetary system, the entire region between the orbit of Neptune and approximately 50 AU would have been swept by first-order mean motion resonances. Thus, resonance capture could occur not only for Pluto, but quite generally for other trans-Neptunian small bodies. Some consequences of this evolution for the present-day dynamical structure of the trans-Neptunian region are (i) most of the objects in the region beyond Neptune and up to ~ 50 AU exist in very narrow zones located at orbital resonances with Neptune (particularly the 3:2 and the 2:1 resonances); and (ii) these resonant objects would have significantly large eccentricities. The distribution of objects in the Kuiper Belt as predicted by this theory is presented here.

1. INTRODUCTION

In the widely accepted paradigm for the formation of the solar system, the planets accumulated in a highly dissipative disk of dust and gas orbiting the protosun, and most planets formed in near-circular and nearly coplanar orbits. The outermost planet, Pluto, is an oddity in this scheme; its orbit is highly eccentric ($e = 0.25$) and inclined (17° to the ecliptic). The large eccentricity means that Pluto crosses the orbit of Neptune, and it traverses a very large region of space from just inside the orbit of Neptune at 30 AU to almost 50 AU. Indeed, soon after Pluto was first discovered in 1930, it was realized that the dynamical lifetime of this new planet was short before a close encounter with Neptune radically altered its orbit (Lyttleton 1936). Three decades later, it was found that a dynamical protection mechanism exists that prevents close encounters between Pluto and Neptune: a 120,000 yr orbit integration of the outer planets by Cohen & Hubbard (1965) showed that Pluto is locked in a 3:2 orbital resonance with Neptune which maintains a large longitude separation between the planets at orbit crossing and causes Pluto's perihelion to librate about a center $\pm 90^\circ$ away from Neptune. Since the Cohen and Hubbard work, orbit integrations of increasingly longer times have uncovered several other resonances and near-resonances in Pluto's motion (Williams & Benson 1971; Applegate *et al.* 1986; Sussman & Wisdom 1988; Milani *et al.* 1989). Perhaps the most important of these "weaker" resonances is the "argument-of-perihelion libration" which ensures that at perihelion Pluto is close to its maximum excursion above the mean plane of the solar system; this has the effect of increasing the minimum approach distance between Pluto and Neptune and between Pluto and Uranus than would otherwise be the case. (It is worth emphasizing that the longitude-of-perihelion libration

centered at $\pm 90^\circ$ relative to Neptune and the argument-of-perihelion libration about 90° are two quite distinct phenomena; the former is associated directly with the 3:2 orbital period resonance lock between Neptune and Pluto.) Pluto's orbit is confined in a very narrow region of relative orbital stability near the 3:2 Neptune resonance, a region in phase space that is bounded by highly chaotic orbits. (See Malhotra & Williams 1994 for a recent review of Pluto's orbital dynamics).

There has, of course, been much speculation as to the origin of this extraordinary orbit of Pluto (Lyttleton 1936; Harrington & Flandern 1979; Farinella *et al.* 1979; Dormand & Woolfson 1980; Olsson-Steel 1988; Malhotra 1993a; Levison & Stern 1995), all but one—Malhotra (1993a, hereafter referred to as Paper I)—of these speculations requiring one or more low-probability "catastrophic" events. A popular theme in the earlier scenarios was that Pluto is an escaped satellite of Neptune. On the other hand, studies and modeling of the physical characteristics and composition of Pluto have led to the conclusion that it accumulated in a heliocentric orbit of ice-rich planetesimals in the outer reaches of the Solar Nebula rather than in a circumplanetary disk (McKinnon & Mueller 1988; Tancredi & Fernandez 1991). In consonance with this, the recent theories by Malhotra (1993a) and Levison & Stern (1995) both propose that Pluto formed in a near-circular coplanar heliocentric orbit beyond the orbits of the giant planets, but they differ in the physical and dynamical mechanisms that placed Pluto in its unusual orbit.

In the model proposed by Levison and Stern, the orbital configuration of the planets is taken as observed today, except that a "test Pluto" is placed in an initially low-eccentricity, low-inclination orbit near the 3:2 Neptune resonance. With some fine tuning of initial conditions, orbit integrations show that such an orbit has its eccentricity and

inclination pumped up to values comparable to those of the real Pluto in a time scale of about 10^7 yr. However, the orbit remains chaotic during this evolution; Levison and Stern then propose that Pluto was “knocked” into the stable 3:2 resonance libration region by one or more dissipative collisions with a neighboring small body or bodies.

Malhotra's theory (Paper I) does not invoke any catastrophic collisions, and is possibly compatible with the standard paradigm for planet formation. [The reader is referred to Levy & Lunine (1993) for reviews of planet formation theory.] In this model, an initially low-inclination, nearly circular orbit of Pluto beyond the orbits of the giant planets evolves into its Neptune-crossing but resonance-protected orbit as a result of early dynamical evolution of the outer solar system. The physical causes of this evolution lie in the late stages of planet formation when the gravitational scattering and eventual clearing of remnant planetesimal debris by the giant planets (and the concomitant exchange of energy and angular momentum between the planets and the planetesimals) may have caused a significant evolution of the giant planet orbits (Fernandez & Ip 1984). In particular, Neptune's orbit may have expanded considerably, and its exterior orbital resonances would have swept through a large region of trans-Neptunian space. During this resonance sweeping, Pluto could have been captured into the 3:2 orbital period resonance with Neptune and its eccentricity (as well as inclination—see Sec. 4) would have been pumped up during the subsequent evolution.

The dynamical mechanisms invoked in this theory are quite general, and would apply not only to the evolution of the trans-Neptunian body labeled “Pluto,” but also to any other members of the trans-Neptune region. While this possibility was implicit in Paper I, it is my purpose in the present paper to make explicit the implications and predictions of this “resonance capture theory” of the origin of Pluto's orbit for the present-day architecture of the Solar system beyond Neptune.

That the outermost parts of the Solar system may be populated by primordial icy planetesimals has been conjectured on both theoretical and observational grounds. For example, Kuiper (1951) suggested this on the basis of theoretical considerations of the genesis of the planetary system from the primordial Solar Nebula. Whipple (1964) and Bailey (1983) speculated on a massive comet belt as the source of unexplained perturbations of Neptune's orbit [although this argument must now be discarded as the post-*Voyager* revisions in the planetary ephemeris no longer show any unexplained residuals in Neptune's motion (Standish 1993)]. Hamid *et al.* (1968) analyzed the orbital plane perturbations of comet P/Halley and concluded that any comet belt between 40 and 50 AU has a total mass less than $1M_{\oplus}$. More recently, it has been suggested that the observed short-period comets with orbital periods $\lesssim 20$ yr originate in a belt of low-inclination bodies just beyond the orbit of Neptune, between 35 and 50 AU (Fernandez 1980; Fernandez & Ip 1983). The older hypothesis that short-period comets originate in a population of near-parabolic Oort Cloud comets (which are perturbed into shorter orbits by the giant planets) appears unlikely: Duncan *et al.* (1988), Quinn *et al.* (1990)

have shown that the orbital element distribution of the observed short-period comets is inconsistent with a source in the nearly isotropic Oort Cloud but is compatible with a disk-like source in a trans-Neptune comet belt, which they call the “Kuiper Belt.” A possible member of the Kuiper Belt was first discovered in 1992 at a distance of 41 AU from the Sun (1992 QB₁, reported in Jewitt & Luu 1993), and several additional discoveries have been reported since (Jewitt & Luu 1995).

The dynamical structure of this putative comet population as determined by the long-term (conservative) gravitational perturbations by the planets has been the subject of two recent studies (Levison & Duncan 1993; Holman & Wisdom 1993). These studies assumed a uniformly distributed initial population in near-circular, low-inclination orbits, and the planets in their present orbital configuration; they sought to determine the extent and nature of any orbital instabilities that might operate on billion year time scales to transport those putative comets into planet-crossing orbits. However, any trans-Neptune population of planetesimals was undoubtedly subject to dynamical evolution during the planet formation process, and the initial conditions assumed in the above studies are not necessarily representative of the state of the Kuiper Belt at the end of planet formation, as acknowledged in Holman & Wisdom (1993). Here I discuss the “dynamical sculpting” of the Kuiper Belt that would have occurred due to the early orbital evolution of the outer planets (during the late stages of their formation) as predicted by the “resonance capture theory” for the origin of Pluto's orbit. The results of this study indicate that the Kuiper Belt would have been “sculpted” into a highly nonuniform distribution early in solar system history, and this structure would be largely preserved to the present epoch: the region beyond Neptune's orbit and up to approximately 50 AU heliocentric distance should have most of the primordial small bodies locked in orbital resonances with Neptune, particularly the 3:2 and the 2:1 orbital resonances which are located at semimajor axes of approximately 39.4 and 47.8 AU, respectively.

The rest of this paper is organized as follows. In Sec. 2, I summarize the “resonance capture theory” for the origin of Pluto's orbit. In Secs. 3 and 4, I describe the numerical simulations conducted to determine the implications of this theory for the dynamical structure of the trans-Neptunian solar system. In Sec. 5, I provide a discussion of the results, a comparison with previous theoretical studies of the Kuiper Belt, and some consequences for observational surveys of the outer solar system.

2. RESONANCE CAPTURE THEORY FOR THE ORIGIN OF PLUTO'S ORBIT

The phenomenon of capture into resonance as a result of some slow dissipative forces is common in nature, and there exists a large body of literature devoted to its study. A well-developed solar system example is the formation of orbit-orbit resonances among the satellites of the giant planets by the action of slow tidal dissipation (see Peale 1986 and Malhotra 1994a for reviews). In general, capture into a stable (long-lived) orbit-orbit resonance is possible when the orbits

of two bodies approach each other as a result of the action of some dissipative process. In Paper I, I proposed that Pluto may have been captured into the 3:2 resonance with Neptune during the late stages of planet formation, when Neptune's orbit expanded outward as a result of angular momentum exchange with residual planetesimal debris. The physics of this mechanism is summarized below.

The giant planets' gravitational perturbations were instrumental in clearing their interplanetary regions of the residual unaccreted planetesimal debris. It is believed that the Oort Cloud—which is a roughly isotropic distribution of comets surrounding the planetary system at distances in excess of $\sim 10^4$ AU—was populated by icy planetesimals scattered outward from the vicinity of the giant planets; estimates of the total mass of the Oort Cloud are in the range $10\text{--}10^2 M_\oplus$, based upon observations of long period comets and extensive theoretical modeling (cf. Weissman 1990). While the formation and dynamical evolution of the Oort Cloud has been a subject of extensive research (indeed, it is subspecialty within planetary science), relatively little attention has been given to the back reaction on the planets themselves of the planetesimal scattering process that populated the Oort Cloud.

Consider the scattering of a planetesimal of mass m_c by a planet of mass M at orbital radius a . If the planetesimal is initially in a near-circular orbit similar to that of the planet and is ejected to a solar system escape orbit, it follows from conservation of angular momentum that the planet suffers a loss of orbital angular momentum and a corresponding change of orbital radius, δa , given by¹

$$-\frac{\delta a}{a} \approx \frac{m_c}{M}. \quad (1)$$

For a planetesimal scattered outward, but one that does not achieve a solar system escape orbit (remaining bound, for example, in the Oort Cloud), the numerical coefficient on the right-hand side of Eq. (1) would be slightly smaller than unity. Conversely, planetesimals scattered inward would cause an increase of orbital radius and angular momentum of the planet. A single, massive planet scattering a population of planetesimals in near-circular orbits in the vicinity of its own orbit would, to first order, suffer no net change of orbital radius as it scatters approximately equal numbers of planetesimals inward and outward.

However, the four Jovian planets acting together evolve differently from this simple picture, as first pointed out by Fernandez & Ip (1984) who modeled the late stages of accretion of planetesimals ("proto-comets") by the proto-giant planets (and the concomitant exchange of energy and angular momentum between the planetesimals and the planets). Their numerical simulations showed a small decrease in orbital radius for Jupiter and significant increases in orbital radius for Saturn, Uranus, and Neptune. The reason for this orbital evolution—in particular that of Neptune—may be understood by means of the following heuristic picture of the

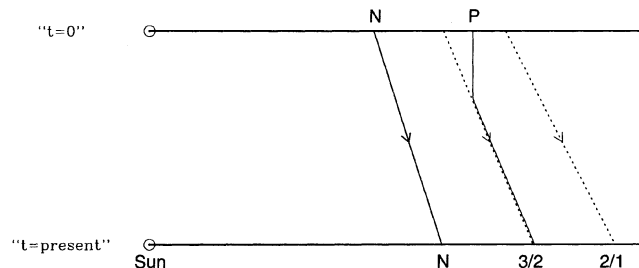


FIG. 1. A schematic diagram to illustrate the outward radial migration of Neptune and its exterior orbital resonances during the late stages of planet formation. The distance from the Sun is along the horizontal direction. Neptune's outward orbital migration is shown along the path marked N–N. For clarity, only two first-order resonances (3:2 and 2:1) are shown (dotted lines). A "Pluto" in an initially circular, nonresonant orbit beyond Neptune could have been captured into the 3:2 resonance and would evolve along the solid line path indicated by P–3/2.

clearing of a planetesimal swarm from the vicinity of Neptune. Suppose that the mean specific angular momentum of the swarm is initially equal to that of Neptune. At first, a small fraction of planetesimals is accreted as a result of physical collisions, and of the remaining, there are approximately equal numbers of inward and outward scatterings. To first order, these cause no net change in Neptune's orbit. However, the subsequent fate of the inward and outward scattered planetesimals is not symmetrical. Most of the inwardly scattered objects enter the zones of influence of the inner Jovian planets (Uranus, Saturn, and Jupiter). Of those objects scattered outward by Neptune, some are lifted into wide, Oort Cloud orbits while others return to be accreted or rescattered; a fraction of the latter is again (re)scattered inward where the inner Jovian planets control the dynamics. In particular the massive Jupiter is very effective in causing a systematic loss of planetesimal mass by ejection into solar system escape orbits. Therefore, as Jupiter preferentially removes the inward scattered Neptune planetesimals, the planetesimal population encountering Neptune at later times is increasingly biased toward objects with specific angular momentum (and energy) larger than Neptune's. Encounters with this planetesimal population produce effectively a negative drag on Neptune which results in Neptune experiencing a net gain of angular momentum and energy, hence an increase in its orbital radius. Note that Jupiter is, in effect, the source of this angular momentum and energy; however, owing to its much larger mass, its orbital radius decreases by only a small amount.

If the above phenomenon did occur in the late stages of planet formation, one consequence of Neptune's orbital expansion is that its orbital resonances would have swept across a range of heliocentric distances comparable to its radial migration (see Fig. 1). During this "resonance sweeping," a small body such as Pluto, initially in a near-circular orbit beyond Neptune, could have been captured and locked into an orbital resonance.

Resonance capture is a complicated dynamical process and a very active subject of research in the nonlinear dynamics literature (see, for example, Wiggins 1991). The transition from a nonresonant to a resonant orbit depends sensitively upon initial conditions, and in the context of the solar

¹The expression given in Eq. (1) in Malhotra (1993a) is incorrect in that the coefficient on the right-hand side of the equation was underestimated by about 20%; I am indebted to S. Tremaine for calling my attention to this.

system, upon the nature of the resonant as well as nonresonant gravitational perturbations, and the rate of orbit evolution due to the dissipative effects. Under certain idealized conditions ("single resonance"), and in the limit of slow "adiabatic" orbit evolution, the probability of resonance capture is relatively straightforward to calculate (Henrard & Lemaître 1983; Borderies & Goldreich 1984). Such a calculation shows that the capture probability for the 3:2 Neptune resonance is 100% for initial eccentricity (before the resonance encounter) less than ~ 0.03 ; the capture probability decreases monotonically for higher initial eccentricities: it is less than 10% for initial eccentricities exceeding 0.15.

Once an object was captured into an orbital resonance, perturbations from Neptune would transfer sufficient angular momentum to it to maintain it in the resonance by expanding its orbit in concert with that of Neptune. A byproduct of this evolution would have been the rapid excitation of the object's orbital eccentricity. This is most readily seen in the following simplified analysis of the first-order perturbations of Neptune on a test-particle orbit. Close to an exterior $j+1:j$ orbital resonance, the first-order perturbation equations for the mean motion, n , and eccentricity, e , of the particle are (cf. Brouwer & Clemence 1961)

$$\begin{aligned}\dot{n} &= 3(j+1)\mu_N n^2 e f(\alpha) \sin \phi, \\ \dot{e} &= -\mu_N n f(\alpha) \sin \phi,\end{aligned}\quad (2)$$

where $\mu_N = m_N / \mathcal{M}_\odot$ is the mass of Neptune relative to the Sun, $\alpha = a_N / a < 1$ is the ratio of semimajor axes of Neptune and the test particle, $f(\alpha)$ is a positive function that can be expressed in terms of Laplace coefficients. $\phi = (j+1)\lambda - j\lambda_N - \varpi$ is the critical resonance angle, with λ and ϖ the mean longitude and the longitude of perihelion, respectively, of the test particle, and λ_N the mean longitude of Neptune.

If the test particle is captured into resonance, its mean motion becomes locked to that of Neptune, so that the following conditions hold:

$$\begin{aligned}(j+1)n &\approx jn_N, \\ (j+1)\langle \dot{n} \rangle &\approx j\langle \dot{n}_N \rangle,\end{aligned}\quad (3)$$

where $\langle \dot{n}_N \rangle$ is the rate of change of Neptune's mean motion as its orbit expands. It then follows from Eqs. (2) and (3) that

$$\left\langle \frac{de^2}{dt} \right\rangle \approx -\frac{2}{3(j+1)} \left\langle \frac{\dot{n}_N}{n_N} \right\rangle = \frac{1}{(j+1)} \left\langle \frac{\dot{a}_N}{a_N} \right\rangle, \quad (4)$$

where the last equality follows from the Keplerian relation between mean motion and semimajor axis ($n^2 a^3 = \text{const}$). Therefore, upon capture into resonance, the test particle's eccentricity is pumped up at a rate determined by the average rate of expansion of Neptune's orbit. The previous equation can be integrated to yield

$$e_{\text{final}}^2 \approx e_{\text{initial}}^2 + \frac{1}{j+1} \ln \frac{a_{N,\text{final}}}{a_{N,\text{initial}}}, \quad (5)$$

where $a_{N,\text{initial}}$ refers to the value of Neptune's semimajor axis at the point of resonance capture. Note that for initially near-circular orbits, the final eccentricity is insensitive to the

initial eccentricity and depends only upon the extent of orbital migration of Neptune.

Applying this result to Pluto, one concludes that if an initially near-circular Pluto was captured into the 3:2 resonance ($j=2$) with Neptune and its eccentricity was resonantly pumped up to its current value of 0.25 in the subsequent evolution, then Neptune was at $a_N \approx 25$ AU at the point of resonance capture. Pluto's initial orbital radius can then be inferred to have been near 33 AU. The 5 AU outward radial migration inferred for Neptune must be regarded as a lower limit, as $a_N \approx 25$ AU is the stage at which Pluto is inferred to have been captured in the 3:2 Neptune resonance. Other trans-Neptune objects captured in the 3:2 resonance at earlier or later times would have their eccentricities pumped up to larger or smaller values, respectively, than that of Pluto. Other first-order resonances of importance in trapping trans-Neptune objects are the 4:3 and the 2:1, currently located at semimajor axes of 36.5 and 47.9 AU, respectively. The 5:3 second-order resonance located at 42.3 AU also has a significant capture probability, and capture is also possible in other higher-order resonances such as the 7:5, and 7:4 located at 37.7 and 43.7 AU, respectively.

3. NUMERICAL MODEL FOR THE EVOLUTION OF THE TRANS-NEPTUNE POPULATION

In the late stages of planetary accumulation, the exact magnitude of the radial migration of the Jovian planets due to their interactions with residual planetesimals is difficult to estimate without a full-scale N -body model. The work of Fernandez & Ip (1984) is suggestive, but simulations of this process stand to profit by the refinements in computational technology that have occurred in the last ten years. For example, the limitations of the software and hardware available to Fernandez & Ip limited the total number of bodies in their simulations to about 2000; therefore, in order to start with a reasonable total initial mass in the planetesimal disk, the masses of their individual planetesimals were in the rather exaggerated range of $(0.02-0.3)\mathcal{M}_\oplus$. Another approximation in their model is the neglect of all but very close encounters between the (proto-) giant planets and the planetesimals. I expect to improve upon this in future work.

For the present, I do not attempt such detailed modeling here, but rather continue to use the simple model outlined in Paper I. Accordingly, the system consists of the Sun and the four Jovian planets with their present masses, together with a population of massless "test particles" representing trans-Neptunian objects. The massive planets perceive their full mutual gravitational interactions and also perturb the test particles, but the latter are noninteracting and do not perturb the planets. (It is perhaps worth emphasizing that there is little overlap between the *trans-Neptunian* population of small objects considered here and the population of residual planetesimals in the *immediate vicinity* of the giant planets whose gravitational interactions with the planets are presumed to be driving the radial migration of those planets.) The orbital migration of the Jovian planets is modeled by a time variation of their semimajor axes according to the following prescription:

$$a(t) = a_f - \Delta a \exp(-t/\tau), \quad (6)$$

where a_f is the semimajor axis at the current epoch and $a_i \equiv a_f - \Delta a$ is the semimajor axis at the starting point ($t=0$) of the simulation. (The epoch " $t=0$ " refers to a time in the late stages of the genesis of the Solar system when the formation of the gas giant planets was largely complete, the Solar Nebula had lost its gaseous component, and the subsequent evolution was dominated by gravitational interactions among the planets and the residual planetesimals. See Levy & Lunine (1993) for details on the various stages of solar system formation.) The orbit evolution given by Eq. (6) was implemented in the equations of motion by means of an additional "drag" force on each planet along the direction, \hat{v} , of the orbital velocity given by

$$\Delta \ddot{\mathbf{r}} = \frac{\hat{v}}{\tau} \left\{ \sqrt{\frac{GM_\odot}{a_f}} - \sqrt{\frac{GM_\odot}{a_i}} \right\} \exp\left(-\frac{t}{\tau}\right). \quad (7)$$

The numerical method described in Malhotra (1994b) was used for the orbit integrations; this method is based upon a second-order symplectic map (see Wisdom & Holman 1991) but modified for additional nongravitational forces. Note that the integration follows the orbits of the four major planets self-consistently (i.e., their mutual gravitational interactions are fully accounted for even as their orbits expand).

4. RESULTS

Here I report the results of several runs based upon the model described above. In all cases, the parameters for the planetary orbit evolution were as follows. The Δa [cf. Eq. (6)] were chosen to be $-0.2, 0.8, 3.0$, and 7.0 AU for Jupiter, Saturn, Uranus, and Neptune, so that the initial semimajor axes of these planets were approximately 5.4, 8.7, 16.3, and 23.2 AU, respectively. These are within the ranges of radial displacements of these planets found in the Fernandez & Ip (1984) calculations. The planet masses and other initial orbital elements were taken from Nobili *et al.* (1989).

The first two runs described here were aimed at determining the current state of a primordial population of small objects beyond Neptune in the Kuiper Belt (up to approximately 50 AU heliocentric distance). In each of these runs, there were 120 test particles—representing the Kuiper Belt objects—with initial semimajor axes distributed uniformly in the range 28–52 AU and all angles (longitude of perihelion, longitude of ascending node, and mean longitude) chosen randomly from the range $(0, 2\pi)$. In Run 1 (a "thin disk"), the initial eccentricities and inclinations were set to 0.01; in Run 2 (a "thicker disk"), the initial eccentricities and inclinations were set to 0.05. The time scale τ for the radial migration of the planets was taken to be 2×10^6 yr. The system was integrated for a period of 2×10^7 yr which is ten times the assumed orbital evolution time scale, τ . At the end of the integration, the final orbits of the planets are very similar to their presently observed orbits. The integration of those test particles that suffered close approaches, i.e., within a Hill sphere radius, with any planet was terminated: the object was presumed to have failed to survive as a Kuiper Belt object. It was found that all objects with initial semima-

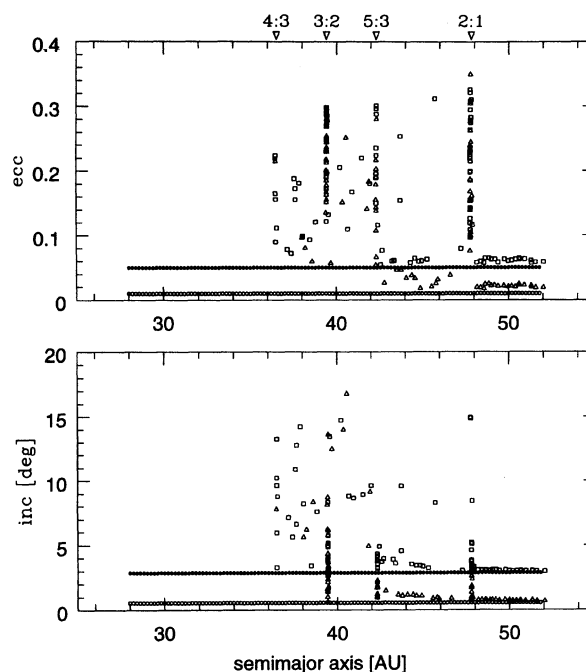


FIG. 2. The initial and final orbits of the surviving Kuiper Belt objects in Run 1 ("thin disk") initial $e=i=0.01$ shows by open circles; final elements shown by triangles; and Run 2 ("thick disk") initial $e=i=0.05$ show by solid circles; final elements shown by squares. For clarity, I show the final semimajor axes averaged over the last two million year period in the runs. The locations of several major orbital resonances with Neptune are indicated at the top of the upper panel. The concentrations of objects at these resonances have significantly large eccentricities and inclinations.

jor axes less than 30 AU and a small fraction of those with initial semimajor axes between 30 and 34 AU failed to survive. The survival rate in the "thicker disk" population was slightly lower than in the "thin disk" (85% vs 91%). (In general, the survival rate can be expected to be sensitive also to the planetary orbital evolution time scale, τ , as discussed below.) Figures 2 and 3 summarize the results obtained in these two runs. Noteworthy features of the surviving test particle population are (i) the final semimajor axes are larger than 36 AU; (ii) the population is highly concentrated at primarily two resonances with Neptune—the 3:2 and the 2:1 resonances located near 39.4 and 47.8 AU, respectively; smaller concentrations are also found at the 5:3, 4:3, 7:5, and 7:4 resonances located at 42.3, 36.5, 37.7, and 43.7 AU, respectively; and (iii) the objects surviving in resonances have significant orbital eccentricities, typically 0.1–0.3 (thus their perihelion distances can be as small as ~ 27 AU).

Three other runs were made with a special focus on Pluto-like orbits captured in the 3:2 orbital resonance with Neptune. In each of these, 120 test particles with initially near-circular, coplanar orbits ($e=i=0.01$) distributed uniformly in $a \in (29, 35)$ AU were integrated for a period of 100 million years with the same model described above. The three runs differed only in the time scale τ of orbit evolution of the planets; τ was set equal to 2, 4 and 10 million years in Run 3, Run 4, and Run 5, respectively. The survival rates in these runs were 79%, 68%, and 53%, respectively. As expected by design, most of the surviving bodies were locked in the 3:2

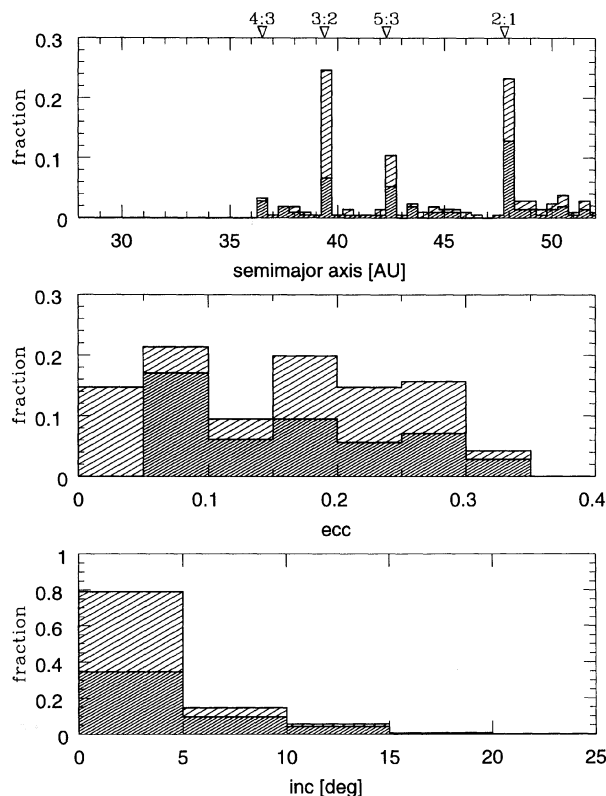


FIG. 3. The distribution of orbital semimajor axes, eccentricities, and inclinations of the surviving Kuiper Belt objects in Runs 1 (lighter shading) and 2 (darker shading). A large fraction of the objects beyond Neptune and up to $a \approx 50$ AU are found in narrow resonance regions, as indicated in the top panel. (The distributions for Runs 1 and 2 are quite similar).

resonance at the end of the evolution. (The exceptions were a few objects which were found locked in other nearby resonances.) An example of the typical orbital evolution of a test particle captured in the 3:2 resonance is shown in Fig. 4 which displays the time variation of those variables that are of particular interest in the dynamics of Pluto's orbit. Observe that the semimajor axis stabilizes at the 3:2 resonance value; the eccentricity and inclination are both amplified; the 3:2 resonance angle, $\phi = 3\lambda - 2\lambda_N - \varpi$ (where λ_N and λ are the mean longitudes of Neptune and the test particle, and ϖ is the longitude of perihelion of the test particle), settles into stable libration about 180° , and the argument of perihelion, ω , also exhibits librations. Long-lived librations of ω about either $+90^\circ$ or -90° (i.e., perihelion near the maximum excursion above or below the mean plane of the solar system) are found in approximately one-third of the objects trapped in the 3:2 Neptune resonance; often, the ω libration center hops between $+90^\circ$ and -90° over 10^7 yr time scales (cf. Fig. 4). The ω librations are strongly correlated with large perturbations of the inclination. (This suggests that the resonance represented by the argument-of-perihelion libration may be responsible for the inclination excitation. However, this conjecture must await further analysis.)

The distributions of the orbital elements of the survivors in these three runs are shown in Figs. 5–7. Noteworthy fea-

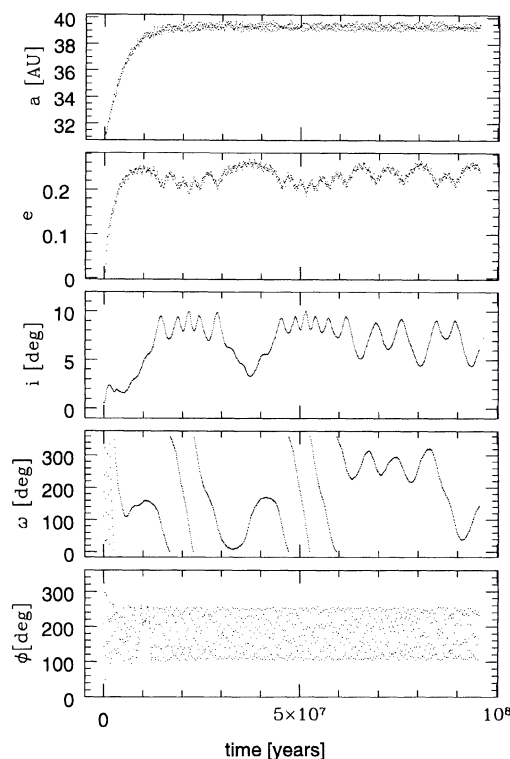


FIG. 4. A typical example of the orbital evolution of a Pluto-like Kuiper Belt object captured in the 3:2 resonance with Neptune. The 3:2 resonance angle, $\phi = 3\lambda - 2\lambda_N - \varpi$, settles into stable librations about 180° . There ensues a secular transfer of angular momentum from Neptune to the object that maintains it in the resonance; the object's orbital semimajor axis increases (in concert with Neptune's) and its eccentricity is pumped up [cf. Eq. (2)]. In many cases, the inclination is also pumped up, but this depends sensitively upon the initial conditions of the orbit; the inclination behavior is highly correlated with the behavior of the argument of perihelion, ω , which often exhibits long periods of libration about either $+90^\circ$ or -90° .

tures are as follows. The eccentricity distribution is virtually identical in the three runs, with the vast majority of the objects having e in the range 0.2 to 0.3 (as expected by design). Most of the objects remain in, relatively low inclination ($<10^\circ$) orbits, but a small fraction (up to 10% in Run 7) have their inclinations pumped up to higher values ($15\text{--}20^\circ$), comparable to that of Pluto. There is a correlation between higher inclinations and larger τ (i.e., slower evolution rate for the planetary orbits).

5. CONCLUSIONS AND DISCUSSION

The unusual properties of Pluto's orbit may be a natural consequence—and a signature—of the early dynamical evolution in the outer solar system. The studies presented here, together with those reported in previous work Malhotra (1993, 1995) provide support for this case. I had shown previously that if Neptune's orbit expanded during the late stages of planet formation, then Pluto could have been captured from an initially near-circular, low inclination nonresonant orbit beyond Neptune, into its current 3:2 resonant orbit; during this evolution, its eccentricity would have been pumped up to the observed Neptune-crossing value while

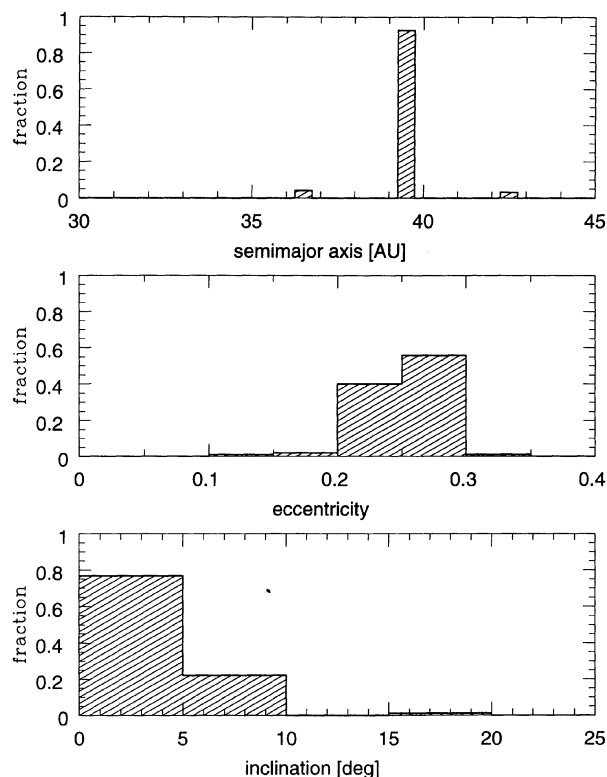


FIG. 5. The final distribution of a , e , and i in Run 3 in which the initial distribution of objects was in near-circular, coplanar orbits in an annulus with a between 29 and 35 AU (designed to yield mostly Pluto-like orbits locked in the 3:2 orbital resonance with Neptune and with e between 0.2 and 0.3). The time scale of orbital migration of the planets in this run was $\tau = 2 \times 10^6$ yr.

maintaining long-term orbital stability. The large inclinations and the argument-of-perihelion librations found in the work presented here augment the previous results and enhance the plausibility of the resonance capture theory for the origin of Pluto's orbit. However, it is also clear that there is a rather small probability for obtaining inclinations as high as Pluto's.

From the observed characteristics of Pluto's orbit, I have shown that it is possible to infer the "initial" locations of both Neptune and Pluto and thus obtain a lower limit of about 5 AU for the magnitude of Neptune's orbital expansion. (With further modeling, one can anticipate that it will be possible to also self-consistently infer the "initial" orbits of the other giant planets.) These inferences are of considerable import for many aspects of planet formation. Some examples are: the provenance of the outer planets may derive from a different and perhaps larger range of heliocentric distance than the immediate vicinity of their current orbits; Jupiter's orbital migration could have influenced the planetesimal dynamics and accumulation process in the region of the asteroid belt; the capture of irregular satellites may have been significantly aided by the radial migration of the (proto-)giant planets. However, a discussion of these points is beyond the scope of the present work.

In this paper, I have followed this "resonance capture"

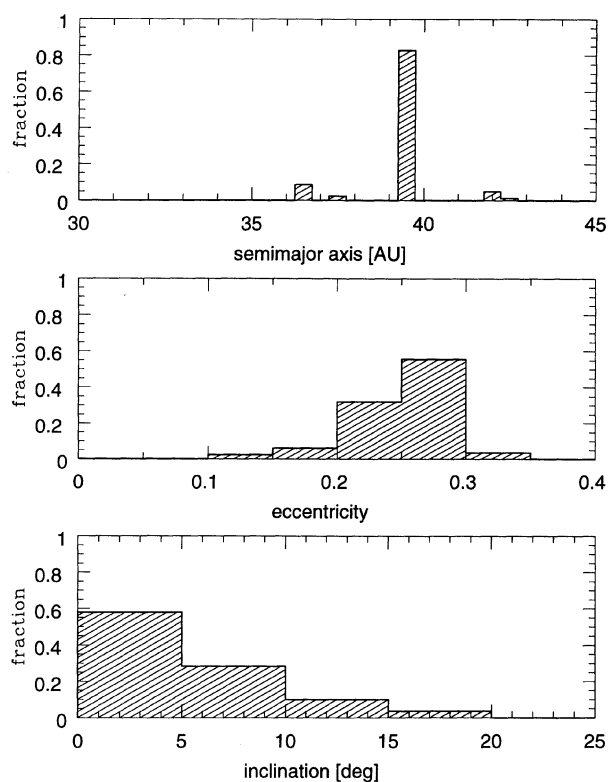


FIG. 6. The same as Fig. 5 but for Run 4 in which the time scale of orbital migration of the planets was $\tau = 4 \times 10^6$ yr.

scenario for the origin of Pluto's orbit to its logical next step in studying its implications for the architecture of the solar system beyond Neptune, i.e., the "Kuiper Belt" of comets approximately between Neptune's orbit and 50 AU. The numerical experiments reported here indicate that the dynamical structure of the Kuiper Belt is dominated by concentrations of objects trapped in orbital resonances with Neptune, particularly at the 3:2 and the 2:1 resonances. These resonant objects move on highly eccentric orbits, with a significant fraction on Neptune-crossing orbits; the inclinations of most of the objects remain low (less than 10°), but a small fraction (up to $\sim 10\%$) are in the 15° – 20° range. Libration of the argument of perihelion (about $\pm 90^\circ$) is not an uncommon occurrence among the resonant objects.

The numerical results also show that the inclination distribution is sensitive to the rate of orbital evolution of the giant planets: longer time scale of the orbit evolution is correlated with higher inclinations (cf. the third panel in Figs. 5–7). The time scales used in the numerical experiments were chosen primarily for computational convenience, and ranged from 2 to 10 million years. As the libration period for the argument-of-perihelion libration (which is correlated with inclination excitation) is several million years, it is clear that the time scale of this resonance is comparable to the time scale of orbital evolution of the giant planets in these numerical experiments. Thus, the sensitivity of the inclination distribution to the rate of orbital evolution of the planets is not surprising as the evolution is not "adiabatic" on this

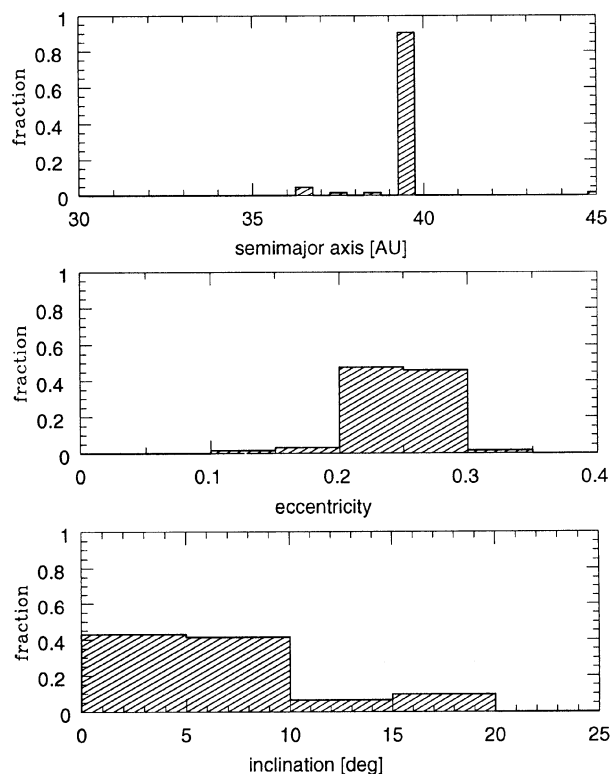


FIG. 7. The same as Figs. 5 and 6, but for Run 7 in which the time scale of orbital migration of the planets was $\tau=1\times10^7$ yr. Note that there is little variation in the eccentricity distributions in Figs. 5–7, but the inclination distribution is progressively broader.

time scale. It is worth noting that the numerical experiments of Fernandez & Ip (1984) indicate orbital evolution time scales of several tens of millions of years, and I expect that more realistic modeling in the future will help determine this time scale better. Comparison of these inclination distributions with observations (when statistically significant numbers of observations of trans-Neptunian objects and their orbital elements are available) may provide a diagnostic of the rate of radial migration of the planets during the late stages of their evolution.

Of course, the integrations reported here were only 2×10^7 – 10^8 yr long, and one might ask if the subsequent nearly conservative evolution over 10^9 yr time scales might not change the profile of the surviving objects. I have extended the orbit integrations to 1 Byr in a few individual cases and have found that the resonance protection remained in place over this longer time period. Although suggestive, this is of course not proof of orbital stability over the 4.5 Byr age of the solar system. A slow leakage of objects out of the orbital resonances over this time scale—either due to collisional evolution, or possibly due to long-term purely gravitational effects, or both—cannot be ruled out, and may be necessary to supply the short-period comet population. In this context, it is also worthwhile to note that resonance capture is not 100% efficient: a small fraction of the original trans-Neptunian population survives the “resonance sweeping” in its primordial nonresonant low-eccentricity, low-

inclination orbits, and may also contribute to the flux of short-period comets. A quantitative study of this point remains to be done.

5.1 Comparison with Previous Theoretical Studies of the Kuiper Belt

There have been two previous theoretical investigations of the dynamics of small bodies in the outer solar system. Holman & Wisdom (1993) and Levison & Duncan (1993) have studied test particle orbital stability on 10^9 yr time scales. For test particles in nearly circular and low-inclination orbits beyond Neptune, these studies found orbital instability on short time scales ($<10^7$ yr) interior to 33–34 AU, an intricate structure of interspersed regions of stability and instability in the semimajor axes range of 34 to 43 AU, and substantially stable orbits beyond 43 AU. The intricacy of the dynamical structure appears to be particularly acute near orbital resonance locations, but as these features are not analyzed in detail in either paper, it is difficult to draw clear conclusions about the relative number density of Kuiper Belt objects that might be expected in these regions at the present epoch. Holman and Wisdom also found substantial “bumps” in the eccentricities and inclinations near the 3:2 resonance and in the eccentricity at the 2:1 resonance, indicating potential sources of planet-crossing cometary orbits in these zones. Both studies indicate that once an object became Neptune crossing, it typically suffered a close encounter with the planet in short order (although a few exceptional cases of long-lived Neptune crossers were detected). In other words, Neptune-crossing orbits do not typically enjoy long-term stability due to resonance protection even when they originate close to resonance locations. This is not surprising, for the stable orbits near Neptune resonances exist only in exceedingly narrow zones in phase space, and a random sampling of initial conditions would have low probability of hitting these zones.

The resonance sweeping mechanism studied in the present paper predicts the inner edge of stability at approximately 36 AU in the semimajor axis, and clear and strong enhancements in the stability and relative population density of objects in orbital resonances with Neptune. In effect, the extensive nonresonant regions of space are depleted of objects as those objects are swept into resonant orbits. In spite of the low phase space volume represented by the resonance libration regions, the dissipative nature of the dynamical evolution in this model very strongly enhances the occupancy rate of the resonance zones. This model predicts somewhat greater “dynamical erosion” in the Kuiper Belt than the above-mentioned studies of Holman and Wisdom and Levison and Duncan, particularly in the regions in between the resonances. The high eccentricities of the resonance-trapped objects make a large fraction of them Neptune crossing yet dynamically long-lived.

5.2 Implications For Observational Surveys of the Kuiper Belt

The chief dynamical feature of the resonance-locked orbits is that the longitude of perihelion *avoids* Neptune. (This

TABLE 1. The first column is magnitude, m ; $s(\text{peri})$ and $s(\text{aph})$ are the minimum radii (in km) of objects of magnitude less than m at perihelion and aphelion, respectively, in orbits of eccentricity 0.2. Columns 2 and 3 are for objects in the 3:2 Neptune resonance (semimajor axis = 39.4 AU); columns 4 and 5 are for objects in the 2:1 Neptune resonance (semimajor axis = 47.8 AU). We assume a mean geometric albedo $A = 0.1$.

m	3:2 resonance		2:1 resonance	
	$s(\text{peri})$	$s(\text{aph})$	$s(\text{peri})$	$s(\text{aph})$
17	690	1570	1020	2324
20	173	395	256	584
22	69	157	102	232
25	17	39	26	58

fact can be inferred from the librations of the resonance angle.) For example, in the case of the 3:2 resonance, the resonance angle librates about 180° , and its maximum libration amplitude (for stable orbits) is about 90° ; the longitude of perihelion librates about a center $\pm 90^\circ$ away from Neptune's mean longitude. The situation is more complicated for the 2:1 resonance, as the location of the center of libration is a strong function of the orbital eccentricity; the center of the perihelion oscillation (relative to Neptune's mean longitude) is at 180° for very small eccentricities; for larger eccentricities (exceeding ~ 0.03), the libration center bifurcates and the two new libration centers drift away (symmetrically) from 180° to nearly $+45^\circ$ and -45° for eccentricities near 0.3. These properties of the resonant orbits will be discussed in further detail in a future publication.

If most Kuiper Belt objects do indeed exist primarily in two resonance bands (the 3:2 and the 2:1), are there particular observational strategies that might improve the discovery rate of these objects? As the results in Sec. 4 indicate, the inclinations of these objects do not exceed 20° (more than 80% of them remain less than 10°). Therefore, sky surveys within $\pm 20^\circ$ of the ecliptic are likely to yield most of the detectable objects.² In a magnitude-limited sky survey, the minimum size of a detectable outer solar system object will vary with its distance and its albedo. Assuming that observations are made at opposition (geocentric distance = heliocentric distance - 1 AU), the size-magnitude-heliocentric distance relationship is given by

$$s = 1803(A/0.1)^{-1/2}r(r-1)10^{-0.2m}, \quad (8)$$

where s is the object radius in km, A is the geometric albedo, r is the heliocentric distance in AU, and m is the magnitude. For a typical orbital eccentricity of ~ 0.2 , the perihelion to aphelion distance varies from 31 to 47 AU in the 3:2 resonance, and from 38 to 58 AU in the 2:1 resonance. The minimum radii of detectable objects at perihelion and aphelion in these resonant orbits are listed in Table 1 for several limiting magnitudes. For a given magnitude and albedo, the minimum radius of a detectable object varies by a factor of

approximately $\sim (1+e)^2/(1-e)^2$ from perihelion to aphelion. For $e=0.2$, this factor is ~ 2.3 . For illustrative purposes, let us assume that the Kuiper Belt objects have a power law size distribution, $n(s)ds \propto s^{-q}ds$, with index $q=2$ (cf. Jewitt & Luu 1995). Then, the number of resonance-trapped objects that are potentially detectable at perihelion is a factor ~ 2.3 greater than those detectable at aphelion. Recall that the perihelia of the 3:2 resonant objects librate near $\pm 90^\circ$ from Neptune. The 2:1 resonant objects (with $e \sim 0.2$) have perihelia librating near $+60^\circ$ or -60° from Neptune. Assuming that, at any given epoch, the resonant objects are uniformly distributed in ecliptic longitude, those objects at longitudes defined by the two quadrants $(45^\circ, 135^\circ)$ and $(-135^\circ, -45^\circ)$ relative to Neptune would be at or close to perihelion. Therefore, without going through a detailed statistical analysis (to take account of factors such as the distributions in eccentricity and resonance libration amplitude), a crude estimate is that discovery rates in these regions of the sky would be greater by a factor of about 2 compared to the two quadrants outside these ranges of longitude. With Neptune currently near 290° ecliptic longitude, the favored quadrants correspond to ecliptic longitude ranges $(335^\circ, 65^\circ)$ and $(155^\circ, 245^\circ)$.

It is interesting to consider the detectability of a Pluto-like object in the 2:1 resonance with Neptune. Such an object (with radius $s \sim 1000$ km, mean geometric albedo $A \sim 0.5$, and orbital eccentricity 0.25) would vary in brightness from $m \approx 15.0$ to $m \approx 17.3$ from perihelion to aphelion. Of all the sky surveys for outer solar system objects to date, Tombaugh's (1961) search is the most likely to have detected such a body. This survey covered all longitudes within approximately $\pm 15^\circ$ of the ecliptic, with a limiting magnitude of about 16.5. Thus, a Pluto-like object in the 2:1 resonance would have been detected if it were at a heliocentric distance less than 49 AU. In a 2:1 resonant orbit (semimajor axis = 47.8 AU) with $e=0.25$, an object spends only about 35% of its orbital period at heliocentric distances less than 49 AU. Therefore, in Tombaugh's survey, there was a $\sim 35\%$ chance of detecting such an object. The more recent surveys for outer solar system objects (Kowal 1989; Luu & Jewitt 1988; Levison & Duncan 1990; Jewitt & Luu 1995) all had limiting magnitudes exceeding 17.3; but they also had much smaller sky coverage; detection probability is much smaller in these surveys for the latter reason. Therefore, a Pluto-like object in the 2:1 resonance cannot yet be ruled out.

Observational surveys of the outer planetary system have recently reported the detection of several objects ~ 100 km in size which are possibly the larger members of the Kuiper Belt (Jewitt & Luu 1995), and it appears likely that even greater numbers of detections will be forthcoming in the near future. As the present work was in progress, Marsden (1994a, b) has reported on the possibility that several of the newly discovered objects may be in Pluto-like orbits, locked in the 3:2 resonance with Neptune. If this is confirmed, it would provide further corroboration for the "resonance sweeping" scenario. It is my hope that the present paper will contribute

²It should be noted that about a third of the resonant trapped objects also exhibit argument-of-perihelion libration; this means that at perihelion (when they are brightest and thus most likely to be detected), such objects would also be near their greatest ecliptic latitude (either above or below the ecliptic); observational searches that are too narrowly confined to the ecliptic may not detect these objects.

to the acceleration of these observational detections and their interpretation within models of the formation and evolution of the solar system.

I am grateful to Scott Tremaine for a critical reading of this paper. This research was done while the author was a

Staff Scientist at the Lunar and Planetary Institute which is operated by the Universities Space Research Association under Contract No. NASW-4574 with the National Aeronautics and Space Administration. This paper is Lunar and Planetary Institute Contribution No. 860.

REFERENCES

- Applegate, J. H., Douglas, M. R., Gursel, Y., Sussman, G., & Wisdom, J. 1986, *AJ*, 92, 176
- Bailey, M. E. 1983, *Nature*, 302, 399
- Borderies, N., & Goldreich, P. 1984, *Celest. Mech.*, 32, 127
- Brouwer, D., & Clemence, G. M. 1961, *Methods of Celestial Mechanics* (Academic, New York)
- Burns, J. A., Schaffer, L. E., Greenberg, R. J., & Showalter, M. R. 1985, *Nature*, 337, 340
- Cohen, C. J., & Hubbard, E. C. 1965, *AJ*, 70, 10
- Dormand, J. R., & Woolfson, M. M. 1980, *MNRAS*, 193, 171
- Duncan, M., Quinn, T., & Tremaine, S. 1988, *ApJ*, 328, L69
- Farinella, P., Milani, A., Nobili, A. M., & Valsecchi, G. B. 1979, *Moon & Planets* 20, 415
- Fernandez, J. A. 1980, *MNRAS*, 192, 481
- Fernandez, J. A., & Ip, W. H. 1983, *Icarus*, 54, 377
- Fernandez, J. A., & Ip, W. H. 1984, *Icarus*, 58, 109
- Hamid, S. E., Marsden, B. G., & Whipple, F. L. 1968, *AJ*, 73, 727
- Harrington, R. S., & van Flandern, T. C. 1979, *Icarus*, 39, 131
- Henrard, J., & Lemaître, A. 1983, *Celest. Mech.*, 30, 197
- Holman, M. J., & Wisdom, J. 1993, *AJ*, 105, 1987
- Jewitt, D. C., & Luu, J. X. 1993, *Nature*, 362 (6422), 730
- Jewitt, D. C., & Luu, J. X. 1995, *AJ*, 109, 1867
- Kowal, C. T. 1989, *Icarus*, 77, 118
- Kuiper, G. P. 1951, in *Astrophysics: A Topical Symposium*, edited by J. A. Hynek, (McGraw-Hill, New York), p. 357
- Levison, H. F., & Duncan, M. J. 1990, *AJ*, 100, 1669
- Levison, H. F., & Duncan, M. J. 1993, *ApJ*, 406, L35
- Levison, H. F., & Stern, S. A. 1995, *Icarus* (in press)
- Levy, E. H., & Lunine, J. I. 1993, in *Protostars & Planets III* (The University of Arizona Press, Tucson)
- Luu, J. X., & Jewitt, D. 1988, *AJ*, 95, 1256
- Lyttleton, R. A. 1936, *MNRAS*, 97, 108
- Malhotra, R. 1993a, *Nature*, 365, 819 (Paper I)
- Malhotra, R. 1994a, *Physica D*, 77, 289 (special issue on Modeling the Forces of Nature)
- Malhotra, R. 1994b, *Celest. Mech. Dyn. Astron.*, 60, 373
- Malhotra, R. 1995, *LPSC*, XXVI, 887
- Malhotra, R., & Williams, J. G. 1994, in *Pluto & Charon*, edited by D. Tholen and S. A. Stern (University of Arizona Press, Tucson)
- Marsden, B. G. 1994a, *IAU Circ.*, No. 5983
- Marsden, B. G. 1994b, *IAU Circ.*, No. 6076
- McKinnon, W. B., & Mueller, S. 1988, *Nature*, 335, 240
- Milani, A., Nobili, A. M., & Carpino, M. 1989, *Icarus*, 82, 200
- Nobili, A. M., Milani, A., & Carpino, M. 1989, *A&A*, 210, 313
- Olsson-Steel, D. I. 1988, *A&A*, 195, 327
- Peale, S. J. 1986, in *Satellites*, edited by J. Burns and M. Matthews (University of Arizona Press, Tucson), pp. 159
- Quinn, T. R., Tremaine, S., & Duncan, M. J. 1990, *ApJ*, 355, 667
- Standish, E. M. 1993, *AJ*, 105, 2000
- Sussman, G. J., & Wisdom, J. 1988, *Science*, 241, 433
- Tancredi, G., & Fernandez, J. A. 1991, *Icarus*, 93, 298
- Tombaugh, C. W. 1961, in *Planets and Satellites*, edited by G. P. Kuiper and B. M. Middlehurst (University of Chicago Press, Chicago), p. 12
- Weissman, P. R. 1990, *Nature*, 344, 825
- Whipple, F. L. 1964, *Proc. Natl. Acad. Sci.* 51, 711
- Wiggins, S. 1991, *Chaotic Transport in Dynamical Systems* (Springer, New York)
- Williams, J. G., & Benson, G. S. 1971, *AJ*, 76, 167
- Wisdom, J., & Holman, M. J. 1991, *AJ*, 102, 1528